Research Centre for Biosystems, Land Use and Nutrition
Institute of Plant Breeding and Agronomy I
Department of Plant Breeding

Genomics of abiotic stress responses and adaptation in sorghum (*Sorghum bicolor* (L.) Moench)

Inaugural Dissertation for a Doctorate Degree in Agricultural Sciences in the Faculty of Agricultural Sciences, Nutritional Sciences and Environmental Management

Examiners

1. Prof. Dr. Dr. h.c. Wolfgang Friedt
2. Prof. Dr. Matthias Frisch

Submitted by
Wubishet Abebe Bekele
from
Addis Ababa, Ethiopia

Giessen 2013

Research Centre for Biosystems, Land Use and Nutrition
Institute of Plant Breeding and Agronomy I
Department of Plant Breeding

Genomics of abiotic stress responses and adaptation in sorghum (*Sorghum bicolor* (L.) Moench)

Inaugural Dissertation for a Doctorate Degree in Agricultural Sciences in the Faculty of Agricultural Sciences, Nutritional Sciences and Environmental Management

Examiners

1. Prof. Dr. Dr. h.c. Wolfgang Friedt
2. Prof. Dr. Matthias Frisch

Submitted by
Wubishet Abebe Bekele
from
Addis Ababa, Ethiopia

Giessen 2013

Plant science has never been more important. The growing and increasingly prosperous human population needs abundant safe nutritious food, shelter, clothes, fibre, and renewable energy, and needs to address the problems generated by climate change, while preserving habitats. These global challenges can only be met in the context of a strong fundamental understanding of plant biology and ecology and translation of this knowledge into field-based solutions.

Grierson et al. (2011) One hundred questions facing plant science research. *New Phytologist* 192, 6-12.

DEDICATION

This work is dedicated to my ever persevering and caring father, Abebe Bekele (Abi); and in loving memory of my late mother, Etaferahu Mulugeta (Emaye).

Table of Contents

1 General Introduction

1.1 Sorghum as a food, feed and bioenergy feedstock

The global food price peaks in 2008 and 2011 are reminders that meeting global demand for food – in the face of climate change, population increase, increased consumption of animal products and growing reliance on biofuels – requires a broader resource base (Ray et al. 2013; McCouch et al. 2013). The world population is estimated to reach 9.1 billion in 2050, causing a rise in global food demand by 50-70% (FAO Director-General, 2009: "How to feed the world in 2050"). This requires interdisciplinary research approaches ranging from bio-physical to socio-economical sciences. Holistic approaches are vital to identify and best integrate alternative solutions. One possible solution is increasing primary productivity *per se*. There are two major types of photosynthetic pathways in plants, known respectively as the C3 and C4 metabolism. Since the most productive natural vegetation and crop plants have C4 photosynthesis, engineering C4 metabolism into C3 cereal crops such as rice and wheat has been proposed as a means to increase productivity of major cereals. However, the differences in leaf anatomy and physiology associated with these two different metabolic pathways make this a far-fetched goal (Hibberd et al. 2008). Increasing the production areas and the productivity of multipurpose C4 crops like maize and sorghum, coupled with a paradigm shift in breeding to exploit the potential of untapped genetic resources and traits, might be a more realistic and productive option (McCouch et al. 2013). The so-called "Green revolution" in the 1950s focused on breeding shorter plants and increasing the harvest index, i.e. the proportion of the grain as compared to the whole above-ground biomass, in major cereals. However, the increase in harvest index in wheat, for example, has reached a plateau at around 60% (Araus et al. 2008). Hence, breeding to improve untapped traits such as root architecture (Hammer et al. 2009) and other physiological traits has been suggested as a means to further improve productivity. Diversification of the number of crops, coupled with a synchronised deployment of classical and biotechnology-assisted breeding, are essential for development of crops and varieties that can provide sustainably higher food, feed and fuel yields in the face of projected climatic changes, such as increases in temperature (2°C warmer), carbon dioxide (550 ppm higher) and ozone concentration (60 ppb) (Jaggard et al. 2010).

Sorghum (*Sorghum bicolor* (L.) Moench), the world's 5th most important cereal crop, is produced for food, feed and bioenergy feedstock in a wide geographic area ranging in latitudes from the tropics to 45° north and south of the equator. Its wide agro-climatic adaptation results from its high phenotypic diversity, which enabled its expansion from West Africa, via the dryland areas of the Sahel into temperate regions of China and Central America. The crop has expanded all over the world as a "failsafe crop" in times of drought

and erratic rainfall (Paterson 2008). Hence, most of the production of sorghum is concentrated in drier and hotter parts of the world, where maize production is difficult or not feasible. The production range covers regions as diverse as the southern USA, northeast Africa and south Asia. The majority of sorghum producers globally are small-scale farmers in drier regions of Africa and Asia, who grow sorghum as a staple food, for feed, for construction material and as a fuel source. In 2012 alone, 37.9 million ha of land was allocated for sorghum production; of this 22.6 million ha (60%) was in Africa (FAO STAT 2013) (http://faostat.fao.org). On the other hand, the country with the largest growing area for sorghum was India (6.3 million ha). The country with the greatest production of sorghum in 2012 was Mexico, followed by Nigeria, the USA and India. In the USA, sorghum is the second most important grain-based bio-ethanol crop after maize, owing to its low input requirements, drought tolerance and cheaper grain price in comparison to maize (Dahlberg et al. 2011).

In Africa and Asia grain sorghum is mainly used as food (55%) in the form of flat breads and porridges. However, there is also a growing interest in production of sweet or biomass sorghum forms to substitute maize as a feedstock for biogas plants in Europe or to produce stem juice molasses or bioethanol in the US (Shiringani et al. 2010; Cai et al. 2013). As their name indicates, sweet sorghum types have a higher sugar content in their stems, which makes them (like sugarcane) amenable to fermentation for bio-ethanol production. However, the fast degradation of sweet sorghum sugar compared to sugarcane calls for research on breeding, genetics and processing to use sweet sorghum as a bioethanol feedstock, just as sweet maize breeding using mutations in polymer biosynthesis enabled to overcome the fast polymerization of free sugars to starch (Lawrence and Walbot 2007). Stefaniaka et al. (2012) praised sorghum as one of the most suitable crops for bioenergy, because of the fact that the starch from the grains can be used for bio-ethanol, the stem and other biomass for biogas, or the stem juice for bio-ethanol; this flexibility is not provided by any other bioenergy crop. In Europe, 35.2 thousand ha of land was allocated for sorghum production in 2012. There is a growing interest in biomass sorghum as a low-input and sustainable alternative to bioenergy maize production in Europe. The complete absence of infestation by the western rootworm (*Diabrotica vergifera*), a serious threat to Europe's maize production (Dahlberg 2011; Shiringani et al. 2010; Shiringani and Friedt 2011; Fiedler et al. 2012), is a particularly valuable characteristic.

Sorghum originated in the dry regions of northeast Africa and is believed to have been domesticated some 8-10 thousand years ago. Archaeological evidence from Egypt identified

seeds showing similar physical and chemical properties to today's cultivated sorghum. Carbon dating on these seeds showed that they were from the time around 8000 before present (Dahlberg and Wasylikowa 1996). Sorghum domestication is believed to have occurred in what is today Sudan and Ethiopia; its distribution by farmers resulted in disruptive selection for adaption to different agro-climatic conditions (Dillon et al. 2007).

The earliest records of sorghum production in Europe date back to 1204, when broomcorn types were described in the Piedmont region of Italy (Dahlberg et al. 2011). Wide use of broomcorn was reported in the17th century in many other Mediterranean countries, and also in central European countries including Germany. However, the relative lack of research and breeding in sorghum compared to other cereals has slowed its expansion. In contrast, maize production in Germany has increased from less than 100 thousand ha in 1960 to more than 2.5 million ha in 2012. Today this makes maize the number one summer crop in Germany (Deutsches Maiskomitee e.V) (http://www.maiskomitee.de/web/public/Fakten.aspx/Statistik). This increase is partly due to long term breeding programs that enabled maize to grow in the cool spring conditions of Germany (Strigens et al. 2013).

1.2 Biology of sorghum: The camel crop

Sorghum belongs to the family of Poaceae (grasses) and the tribe Andropogoneae. This tribe consists of 85 to 90 genera including the genus sorghum. Of six subgeneric groups, cultivated sorghum and its wild progenitors are among six species classified as Eusorghum (Dillon et al. 2007; Kellog 2013). Cultivated sorghum derived from crosses between *Sorghum drummondi* and *Sorghum verticifollium*, which occurred in several sub-Saharan African countries between 8 and 10 million years ago (Kellog 2013). The cultivated *Sorghum bicolor* has four wild subspecies and five cultivated races, which are widespread in different parts of Africa and show large differences in inflorescence and panicle structure. These differences enable them to adapt to different local environmental conditions, ranging from western Africa with its high humidity and irregular rainfall patterns, to the Sahel and eastern Africa. The four wild races of *S. bicolor* races are *arundinaceum, virgatum, aethopicum and verticilliflorum*, an in-group assigned to the subspecies *verticilliforum* (formerly *arundinaceum*; Kellog 2013). Though the differences are morphological, a recent phylogenetic study showed that genotypes cluster preferentially according to their geographic origin, followed by clustering into races or morphotypes (Morris et al. 2013; see Figure 1.1).

Sorghum is a diploid organism (2n=20) with a relatively small genome size of 730 Mbp. Phylogenetic studies show that sorghum is closely related to important crops such as maize

and particularly sugarcane. Sugarcane is the single most important bioenergy and sugar crop, and is the crop with the greatest tonnage on Earth. Within the family Poaceae, the split between sorghum and sugarcane occurred only around 5 million years ago (mya), after the split from maize (12 mya) and rice (42 mya) (Paterson 2008). Sorghum is therefore a key to understanding the evolution and domestication history of the most efficient and economically important crop plants. Sorghum is considered to be the best model for C4 crops with an African origin, making it an interesting addition to plant models genomes from other continents like Arabidopsis (temperate zones), maize (America) and rice (Asia). The close relationship of sorghum to economically important polyploid crops such as sugarcane and switchgrass (*Panicum virgatum*) is a good reason to make use of its considerably smaller, less complex, diploid genome (Lawrence and Walbot 2007). The sorghum whole-genome sequence assembly was published by Paterson et al. (2009). One of the interesting findings from analysis of the genome sequence was the presence of greater copy numbers of expansins, Cytochrome P450 and microRNA-169g genes than in the rice, Arabidopsis or poplar genomes. Proliferations of these sequences are thought to be a key factor in the adaptation of sorghum to marginal conditions.

Sorghum is sometimes called a "camel" among crops due to its very high water-use efficiency. One major reason for this is its C4 photosynthesis. C4 plants have a peculiar leaf anatomy, in which the enzyme Ribulose-1, 5-biphosphate-carboxylase/-oxygenase (RuBisCO) is localised in the bundle sheath cells with high $C0_2$ concentrations. This increases the efficiency of photosynthesis by circumventing the tendency of RuBisCO to undergo oxygenase activity. The second characteristic of sorghum that contributes to its hardiness is its deep root system. At early seedling development sorghum, in contrast to maize, has one or two seminal roots that penetrate deep into soil layers (Singh et al. 2010). The root-to-shoot ratio of sorghum is increased due to a relative reduction in shoot biomass; however there are also reports of increased root growth as a result of drought stress (Salih et al. 1999). Recent studies by Schittenhelm and Schroetter (2013) in Germany and France showed that sweet sorghum has superior drought tolerance, attributed to its deeper root penetration and root biomass than maize. Sorghum is also known to have an epicuticular waxy layer that hinders non-stomatal water loss (Premachandra et al. 1994).

Another distinctive feature of sorghum that contributes to its ability to withstand drought stress is its stay-green character. Stay-green or non-senescence is the ability of a plant to retain green stems and a green flag leaf up to 60 days after flowering under water-limiting

conditions (Rosenow 1977; Borrell et al. 2000). In sorghum, several quantitative trait loci (QTL) are known to be involved with the stay-green character, however only the four major stay-green loci on chromosomes Sb02, Sb03 and Sb05 have been investigated in detail (Kebede et al. 2001; Harris et al. 2007).

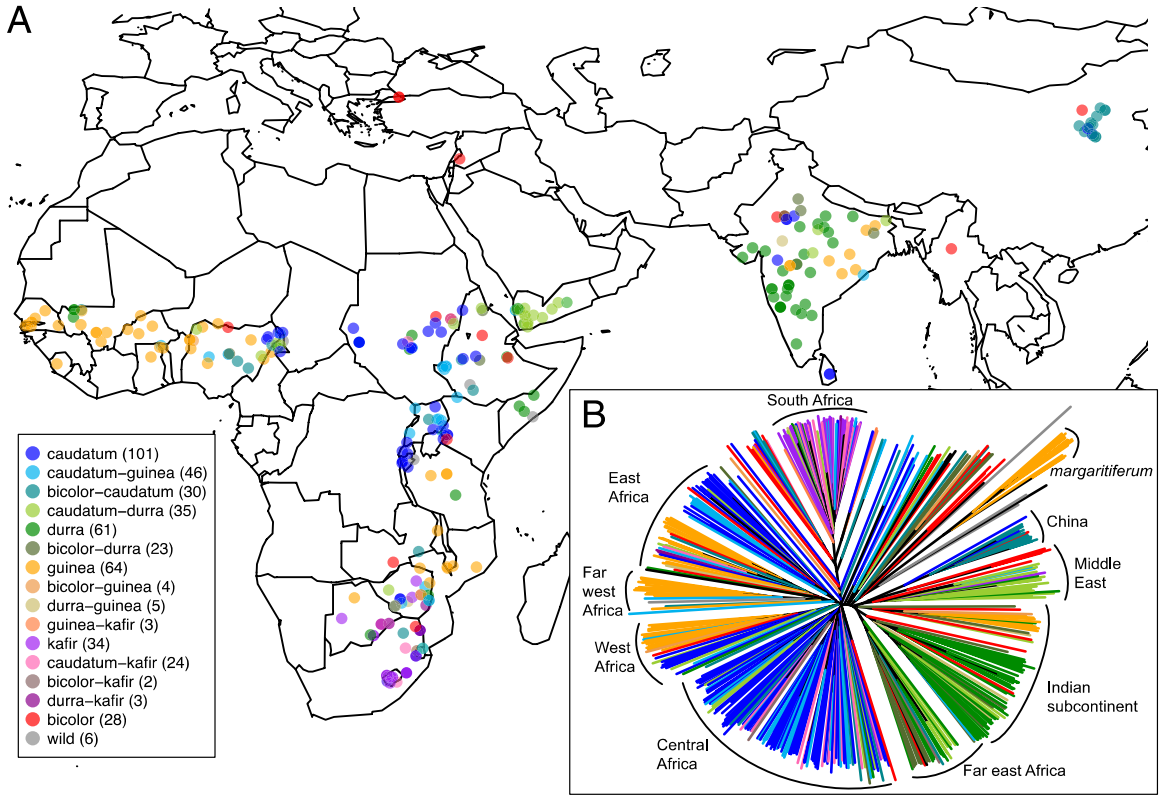


Figure 1.1 Distribution and genetic diversity of sorghum races with known geographic origins. A) Geographic origin of 469 genetically diverse sorghum accessions studied by Morris et al. (2013); B) Neighbour-joining graph of the same 469 lines clustered into morphotypes within a region. Source: Morris et al. (2013).

Sorghum shoots and roots also produce allelopathic compounds, depending on the environment, organ and growth stage, which can suppress the growth and establishment of competing weed species. Crop rotation systems that integrate sorghum, or intercropping or addition of green sorghum manure is therefore widely used to control weed infestation (Weston et al. 2013). The application of sorghum root and shoot extracts for weed control has been tested in several crops. The active ingredients include Sorgenole, phenolic compounds, Dhurrin and cyanide. Sorgenol, one of the first allelopathic compounds identified in sorghum, also plays a role in the initiation of the parasitic plant pest Striga. This parasitic weed affects crops such as sorghum, maize and cowpea on more than 500 million ha of sub-Saharan Africa. Detailed studies on the biochemical interactions of Striga with sorghum compounds

enabled identification of alleles for Sorgenol production in sorghum, subsequently leading to the development of resistant varieties (Ejeta 2007).

As a tropical plant sorghum requires high temperature at different stages of its development. Germination and emergence have a basal temperature around 10°C, while other seedling development processes have a basal temperature around 15°C. The optimum temperature for sorghum growth is between 20-40°C. The other phase of sorghum development with high sensitivity to chilling temperatures is during pollen formation: Temperatures below 15°C are known to cause high pollen sterility (Peacock 1982; Yu and Tuinistra 2001). Field establishment studies on sorghum cold tolerance showed the presence of high genetic variability for germination, emergence and other seedling development processes under field and controlled conditions (Brar et al. 1992; Franks et al. 2006). In these studies, Chinese Kaoliang forms were the main sources of cold tolerance. However, a number of germplasm characterizations showed that sorghum lines from the highlands of Africa (Rwanda and Ethiopia) and Yemen also harbour genetic variability for some of the crucial traits allowing effective establishment of sorghum under low temperature conditions. The biomass yield of sorghum is highly dependent on the time at which the temperature is above the basal temperature between sowing and flowering (Craufurd et al. 1999; Hammer et al. 2010). Hence, breeding for early stage cold tolerance in biomass sorghum is important, so that the crop can benefit from the long growing seasons in the European summers.

Sorghum is a short day plant. A systematic breeding programme in the USA in the 1970s introgressed different mutations for day-length insensitivity (*Ma1-Ma4*) and plant height (*Dw1-Dw4*) into exotic sorghum lines from all over the world (Brown and Paterson 2013). This has enhanced breeding progress by providing sorghum breeders with lines whose flowering behaviour was more amenable for crosses and enabling development of grain sorghum lines that can be harvested by combine harvesters. Additional adaptation loci implemented for breeding purposes include the maturity loci *Ma6* and *Ma7* and the thermo-sensitive allele *T*, a dominant late flowering gene that functions only under night temperatures above 20°C (Tarumoto 2011). Another important finding that accelerated breeding in sorghum was the discovery of a cytoplasmic male sterility system, based on the Kafir nuclear restorer with a Milo cytoplasm system (Stephens and Holland 1954) that allowed targeted exploitation of heterosis. Sorghum is a predominantly self-fertilized crop but has up to 30% outcrossing depending on the panicle structure and environmental conditions. Hence it is

suitable for the development of hybrid cultivars that exploit the yield benefits caused by heterosis.

One fascinating area for sorghum research is understanding the biology of the world's most noxious weed, Johnson grass (*S. halpense*). Johnson grass is a cross between *S. bicolor* and *S. propqunium*, which gave rise to a fertile tetraploid (2n=4x=40) that can propagate with seeds as well as vegetatively. It has considerable variability compared to its progenitors and can also colonize wide geographic regions. These peculiar characters of Johnson grass are today being utilized in functional genomics studies to understand the genetic and functional variability within this "wonder-weed" (Jessup 2013).

Sorghum has become an essential component of grass genome research, revealing interesting examples for conserved but parallel domestication of grasses. One example is the seed shattering gene, which showed parallel domestication in sorghum, maize and rice (Lin et al. 2012). Genome sequencing programs in Miscanthus, maize and sugarcane have all benefited from information gained from the smaller, less complex but closely related sorghum genome. In particular, sweet sorghum is a perfect model for molecular genetics, physiology, biology and breeding of sugarcane (Paterson et al. 2009; Lawrence and Walbot 2007), both crops store sugar in their stems, but sorghum has the great advantage for genetic studies because of a relatively short growing period, a small diploid genome, and the ability to propagate via seeds.

1.3 Abiotic stress responses in plants

Because of their sessile nature, plants need to adapt to abiotic factors that are continuously changing in their habitats. The temporal and spatial variability, the extent and duration of an abiotic factor determines the level of the stress it causes in plants (Gaspar et al. 2002). Annually, abiotic stresses are known to decrease the productivity of natural and agricultural systems on average by more than 50% (Boyer 1982; Qin et al. 2011). Abiotic stress can cause, physical, biochemical and phenological changes that are reflected as tolerance or susceptibility, in extreme cases even leading to death.

Based on their adaptation to habitat heterogeneity there are three different types of plants. The first group are specialist plants that have developed evolutionary adaptations to maximize fitness to a specific environmental condition. Depending on the severity of selection, these types of species might become endemic or locally adapted. The second group is called the generalists, which develop "general purpose" phenotypes that ensure survival in several

environments. At a local level, however, a generalist would have lower fitness in comparison to a specialist species adapted for the specific environment. The third and possibly most successful group are plants showing an adaptive plastic response, in which genotypes can produce an adapted phenotype based on the environmental conditions (Des Marais and Juenger 2010). This strategy deploys environmentally specific alterations of growth and development, such as stomatal closure in response to drought.

One of the most important abiotic stresses influencing both natural and agricultural ecosystems is osmotic stress. Osmotic stress can be caused by drought, salinity or mild coldness; hence these stresses can often be addressed together. Drought is usually the combined effect of low precipitation and high evaporation rate from the soil surface. Salt stress increases the soil water potential near the rhizosphere, making it difficult for the plant to extract water from the soil. Mild cold stress affects the general plant metabolism rate and membrane fluidity, hindering the diffusion of water and other solutes into and out of cells (Des Marais and Juenger 2010).

Osmotic stress in plants causes various phenological changes and disturbances, at several organizational levels ranging from cells to organs. The most common physiological and biochemical changes that affect plant homeostasis and growth are membrane disorganization, excessive reactive oxygen species (ROS) and protein denaturation. The latter effect results in a decline of photosynthetic activity and damage to cellular structures leading to stunted growth, reduced fertility and premature senescence. If the stress affects the physiology of the plant at different organizational levels, the response will also occur at different levels. There are several ways plants respond to osmotic stress, the broadest and most widely used categories being classified as stress avoidance and stress tolerance. Stress avoidance is a mechanism aimed at balancing water uptake and water loss. This can be achieved by avoidance of low water potential using mechanisms such as stomatal closure, reduced shoot area or high root-to-shoot ratio. However, due to the conservative nature of avoidance mechanisms, they carry a cost in terms of yield decline. Stress tolerance mechanisms, on the other hand, come into play when stress avoidance mechanisms are no longer sufficient. Stress tolerance mechanisms are aimed at ensuring continued growth and survival despite low water potential experienced by the plant. One option is dehydration tolerance, which protects cellular damage caused by water loss using ROS scavenging compounds such as super oxide dismutase (SOD) (Cruz de Carvalho 2008). The other main strategy is the implementation of

dehydration avoidance mechanisms such as accumulation of solutes, or cell wall stiffening, to minimize water loss (Krasensky and Jonak 2012; Claeys and Inzè 2013; Verslues et al. 2006).

Stress avoidance mechanisms, which limit growth even after the stress has subsided, can have a huge yield penalty, but at the same time the maintenance of growth under stress conditions can put survival in question. This dilemma was referred to by Claeys and Inzè (2013) as "the agony of choice". The complex interplay between stress survival and growth maintenance has major implications for abiotic stress breeding. Genotypes that are good at surviving extreme drought are usually not the best genotypes under mild stress or optimum conditions. Hence, finding genotypes with sufficient developmental plasticity to cope with changing environmental conditions is quite important.

Temperature is one of the most important abiotic stress factors determining the geographic distribution of any given species. The duration of warm or moist periods are the two most important criteria determining the agroecological zones, the natural vegetation or the length of the growing period. Temperature stress is a general term for suboptimal temperatures, which can be higher or lower than optimal temperatures. Grasses are one of the most widely distributed plant groups, ranging from tropical forest ecosytems to temperate regions. Adaptive evolutionary processes, driven by dynamic genome rearrangements and duplications affecting low-temperature stress response genes, are believed to speed up the rate of mutation and speciation. Such processes gave rise to some of the most important winter crops, such as wheat, rye and barley in the group Pooidae (Vigeland et al. 2013). Furthermore, many temperate plants including some crop plants are tolerant to freezing temperatures (below 0°C), as a result of acclimatization (pre-exposure to chilling temperatures between 5-15°C, resulting in changes to their physiological and metabolic status). In contrast, many tropical crops such as sorghum are considered sensitive to temperatures below 20°C (Peacock 1982; Chinnusamy et al. 2007). Breeding for chilling tolerance in maize, based on an improved understanding of the genetics and physiology of chilling stress responses, has enabled its enormous expansion into colder regions well beyond its tropical origin. The major traits considered for improvement of chilling stress tolerance in maize were seedling vigour and photosynthetic activity (Leipner 2009). To some extent the role of root growth and development were also investigated, under controlled and field conditions (Hund et al. 2004). Several QTL for seedling vigour, chlorophyll content and fluorescence traits associated with stress response were identified (Leipner 2009), and factorial cross phenotypic and expression

studies showed variation in seeds for germination and desiccation tolerance, along with heterosis for early stage chilling tolerance (Kollipara et al. 2002; Bhosale et al. 2007).

As a tropical crop plant, the optimum temperature for sorghum growth is between 20-40°C and a temperature more than 40°C or below 15°C causes a considerable stress in growth, development and yield (Peacock 1982). The temperature minimum at which plants stop growing is called the basal temperature. The basal temperature can be determined for different processes and under various conditions. According to Payne et al. (2003) and Fiedler et al. (2012), emergence of sorghum seedlings ceases at soil temperature of around 8°C or lower, whereas seedling development ceases under a basal temperature of around 12°C in a genotype dependent manner. On the other hand, chilling tolerance at emergence does not make sense if the seedling is unable to tolerate the chilling conditions (Leipner 2009; Knoll et al. 2008). Germplasm with interesting chilling tolerance has been so far identified such as the Chinese Kaoliang (Franks et al. 2006) and other lines from high altitude areas of Africa (Payne et al. 2006) and preliminary QTL analysis undertaken on the field and controlled conditions identified interesting QTL regions (Knoll et al. 2008; Burow et al. 2011). Emergence processes modelling and association mapping on a diversity panel could map and identify genetic regions that are involved in emergence and stability parameters (Fiedler et al. 2012).

Breeding for adaptation and improved abiotic stress has made a major contribution to crop yield increments worldwide. For example, World maize has shown very high yield increases of around 2% per year for the past 5 decades, with 75% of this increase being attributed to genetic improvement. Much of this increment is not due to changes in primary traits such as harvest index or heterosis, rather due to improved abiotic stress adaptation including improved photosynthetic activity during the grain filling stage (stay-green) and an increase in kernel number and better partitioning for kernel number during drought- and heat-sensitive flowering and ripening stages (Borrell and Hammer 2000; Kebede et al. 2001; Blum 2013). Araus et al. (2008) and Hammer et al. (2009) demonstrated that a considerable proportion of the yield increment in US maize hybrids in the past few decades can be explained by changes in root architecture.

1.4 Genetics and breeding of complex traits in the genomics era

The majority of important agricultural traits are quantitative traits or complex traits. The term complex trait refers to phenotypes that are controlled by many small-effect genes (Lander and Schorck 1994) and additional modified by environment. The phenotypic variance of a metric character observed can be partitioned into three major components: the genotypic variance,

the environmental variance and the genotype-by-environment interaction. The heritability of a trait refers to the contribution of the genetic variance in relation to the total phenotypic variance (broad-sense heritability) or the additive genetic variance vs. the phenotypic variance (narrow-sense heritability). The task of any breeder is to find reliable traits that can be used to select genotypes that best fit a target environment (Falconer 1989). The success of a breeding program, i.e. response to selection (R) relies directly on the heritability of the trait, the accessible diversity and the selection intensity as factors determining the value of R.

The application of DNA marker technologies for selection has revolutionized genetic analysis and breeding, since DNA-based markers are basically more stable than many phenotypic traits (Collard et al. 2005). Until relatively recently, a major bottleneck for the application of molecular markers was their comparatively high cost and the low throughput of genotyping. In the past decade DNA sequencing and high-throughput genotyping technologies have made extraordinary advances. Today it is possible to resequence a whole sorghum genome in just a few weeks for only a few thousand dollars, and reference genome sequences have been published for numerous model plants and crops. Among these are more than 10 whole genome sequences of grass species, including rice (Matsumoto et al. 2005), sorghum (Paterson et al. 2009) and maize (Schnable et al. 2009). Cost-effective re-sequencing in crops such as rice, maize and sorghum is giving breeders and geneticists essential tools for comparative genomics, functional genomics, genetics and breeding (Zheng et al. 2011; Nelson et al. 2011; Ganal et al. 2011; Varshney et al. 2013).

Technological advances in life sciences have contributed much to our understanding of genome dynamics and its contribution to any given phenotype. The technology advancement is also leading the way scientists conduct genetic analysis. This has changed the way markers are developed and applied in genetic analysis and breeding. In the past, a major bottleneck in marker-assisted breeding or genetic analysis of complex traits was the lack of sufficient or suitable molecular markers. Today precise and fast phenotyping is often more critical and expensive than getting SNP data for thousands of markers (Ziyomo and Bernardo 2013; Dhont et al. 2013). Chapter 3 of this thesis (Bekele et al. 2013b) reports a study of the patterns of sequence variation in sorghum and identifies genome-wide single-nucleotide polymorphism (SNP) markers for the development of a 3000-SNP array.

One advantage of partially or inbreeding crops, including sorghum, is the ability to create immortal, homozygous populations, such as recombinant inbred line (RIL) or double haploid (DH) populations that can be phenotyped extensively under multiple environments and

conditions. Such data is an essential basis for understanding the inheritance and genetic variance of complex quantitative traits, for example by applying biparental QTL mapping or association mapping techniques. Conventional QTL mapping attempts to find linkage between molecular markers and genomic regions controlling the phenotype, and at the same time to quantify the contribution of that region to the observed phenotypic variance. Conventional QTL mapping in biparental populations has the potential to identify rare alleles and can give a good estimate of the allelic effects in a given genetic background. On the other hand, recombination is limited to the few numbers of genotypes that can be handled in a single QTL experiment (generally a few hundred individuals).

Association mapping uses natural variation within the species that is dependent on historic recombination between ancestors split many generations ago. This greatly increases the resolution for QTL mapping, however, association mapping based on linkage disequilibrium is less effective at estimating the effects of alleles in general and rare alleles specifically, which may nevertheless be valuable for breeding (Collard et al. 2005). Several improvements to these methods have been suggested, such as joint linkage association mapping (JLAM) or nested association mapping (NAM). Such techniques have been widely used in maize, sorghum and a number of other crops (e.g. Yu et al. 2008; Lu et al. 2010; Jordan et al. 2011).

Since a single phenotype has several layers of control from the gene to the phenotype, genetic studies that take into account transcriptomic and metabolomic data can potentially better understand the regulation of quantitative phenotype responses to environmental stress. In one example, Munkvold et al. (2013) reported a systems analysis on transcriptome data from wheat embryos derived from different environments that identified specific gene expression modules responsive to environmental factors. In another recent example, Wang et al. (2013) used systems modelling combined with expression QTL analysis to identify interesting candidate genes and co-regulated gene expression networks influencing rice root development. With careful experimental design, adequate data management and exact analysis, proteomics, metabolomics and transcriptomics data can enable systems biological approaches that integrate multidimensional data to improve understanding of complex phenotypes. This approach is considered as the future of several biological disciplines including plant breeding and genetics (Munkvold et al. 2013; Cramer et al. 2011; Cooper et al. 2009).

However, omics techniques need to be implemented with care. Blum (2011) argued that the genomics of complex traits under drought stress are much more complex than the traits

themselves. As an example he discussed two field-based traits: anthesis silking interval (ASL) in maize and stay-green in sorghum. Despite the fact that these traits are efficiently utilized traits in breeding for drought tolerance, several genomic studies on these traits could not clone the responsible genes and were unable to fully explain their molecular basis. Therefore, to better understand the environment component of phenotypic variance, Blum (2011) recommended the integration of reliable field-based data into research endeavours that use genomics and post genomics to dissect quantitative traits.

1.5 Genomics assisted adaptation of sorghum to Central Europe

Maize (corn) is the number one spring-sown crop in Germany and many other European countries. The increase in maize production area was made possible by its adaptation to European climatic conditions, day length and agronomic systems (Leipner 2009). Adaptation of a new crop to any agricultural system is a challenging task. Soybean introduction to the USA is one excellent example of a successful introduction and adaptation of a crop. The US government invested 5 million US dollars between 1912 and 1942 to introduce the crop from China; in the year 2000 US soybean exports were worth around 6.6 billion dollars (Prohens et al. 2007).

Chapter 2 of this thesis discusses tolerance to chilling stress in early stage seedling development as a prerequisite for the successful establishment of sorghum as an alternative to maize production in Germany. Genomic analyses were implemented to better understand the complex genetics of adaptation of sorghum to central European conditions (Bekele et al. 2013a). Similar retrospective studies in potato revealed alleles that determined the ability of potato to adapt and form tubers in European conditions (Kloosterman et al. 2013). In sorghum, Thurber et al. (2013) used genotyping-by-sequencing of one thousand exotic and converted lines to retrospectively map the genes that were selected during grain sorghum adaptation breeding in the USA.

Biomass and sugar yield in energy sorghum is determined to a great extent by the plant height, the stem diameter and the amount of light intercepted by the plant. High yields are only achievable if the growing season is sufficiently long (Vermerris and Saballos 2013). Since plant height and maturity are easy to assess and have a relatively simple inheritance in sorghum, it is possible to breed for such traits even without molecular markers. However, complex interactions between maturity genes and the temperature locus T need to be considered when such traits are used as selection criteria for adaptation (Tarumoto et al. 2003). Environmentally dependent traits that are more complex and difficult to phenotype,

such as early stage chilling tolerance, can therefore benefit from the application of genomics tools.

Modern biotechnology and genomics technologies, applied in the context of an increased understanding of physiology, can shorten the time required for successful adaptation of a crop to a new agro-ecosystem. Marker assisted selection (MAS) for major genes, on the one hand, or the application of genomic selection (GS) in plant breeding for more complex traits, can increase the genetic gain per selection cycle (Jannink et al. 2010). Sorghum adaptive traits such as emergence and seedling vigour under chilling conditions are controlled by several small effect QTL that interact strongly with the environment (Knoll et al. 2008; Burow et al. 2011; Fiedler et al. 2012). Precise phenotyping and genetic analysis can identify linked markers that can be used for MAS of specific QTL. However, such QTL frequently do not show comparable effects in different genetic backgrounds due to epistatic interactions. This genetic background effect often minimizes the transferability of QTL from experimental mapping populations into elite breeding mateirals (Jannink et al. 2010). One solution for this problem are GS approaches. GS uses genome-wide markers and phenotype data from a "training population" to estimate the effect of each marker on the phenotype. The sum of all marker effects of an individual are used to calculate the genomic estimated breeding values (GEBVs) that predict phenotypes in a candidate population using only marker data. This method is particularly interesting for traits with several small-effects QTL and lower heritability.

GS was developed and widely used by animal breeders (Meuwissen et al. 2001; Hayes and Goddard 2010). In plant breeding GS has been tested in several crops including maize (Riedelsheimer et al. 2012; Crossa et al. 2013), wheat (Poland et al. 2012), sugar beet (Würschum et al. 2013; Hofheinz et al. 2012) and sugarcane (Gouy et al. 2013). Although GS is mainly used to directly select yield or yield components, GS for biotic stress resistance (e.g. Fusarium head blight in barley; Lorenz et al. 2012) and abiotic stress tolerance like drought tolerance or yield improvement under drought stress (Ziyomo and Bernado 2013) and winter survival in European winter wheat (Zhao et al. 2013) have shown good prediction accuracies. The major advantages of GS are the potentially cheaper cost of genotyping than field phenotyping, along with the possibility to reduce the duration of each selection cycle and hence more rapidly advance breeding progress (Ziyomo and Bernando 2013). Given the valuable genome resources available for sorghum, including the genome-wide sequence data and SNP array described in this thesis (Bekele et al. 2013b), GS is set to play an important

role in bioenergy sorghum breeding. The final chapter of the thesis gives examples for the application of the sorghum SNP array for QTL mapping and GS in bioenergy sorghum research and breeding.

1.6 Scope and Aims

The work described in this thesis is part of an effort to understand the early stage chilling tolerance of sorghum, as a prerequisite for the establishment of sorghum in Europe (Chapter 2). The work explores genetic variation of sorghum at a DNA sequence level, in order to develop genomic tools for genetic studies and breeding of essential adaption traits for Europe (Chapter 3). The general discussion (Chapter 4) includes selected examples that show the potential application of genomics-assisted breeding for improvement of sorghum for bioenergy. The work ends with an outlook on potential applications of systems biology in plant breeding.

Considering the huge potential of sorghum and the complex genetics governing its adaptation and yield, this work was set out with the following overall aims:

- Genetic analysis of early stage seedling development of sorghum under chilling temperature conditions;
- Exploration of patterns of genetic variability in sorghum at a DNA sequence level;
- Development of genomic tools for genetic analysis of abiotic stress response and adaptation of sorghum to temperate conditions;
- Testing the feasibility of genomics-assisted breeding of bioenergy sorghum for adaptation and yield.

2 Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions

Bekele WA, Fiedler K, Shiringani A, Schnaubelt D, Windpassinger S,

Uptmoor R, Friedt W, Snowdon RJ

Plant, Cell & Environment 2013

doi: 10.1111/pce.12189. [Epub ahead of print]

Plant, Cell & Environment

Plant, Cell and Environment (2013) doi: 10.1111/pce.12189

Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions

WUBISHET A. BEKELE[1], KARIN FIEDLER[2], AMUKELANI SHIRINGANI[1], DANIEL SCHNAUBELT[1,3], STEFFEN WINDPASSINGER[1], RALF UPTMOOR[4], WOLFGANG FRIEDT[1] & ROD J. SNOWDON[1]

[1]Department of Plant Breeding, Justus Liebig University, Giessen, 35392 Giessen, Germany, [2]Department of Biological Production Systems, Leibniz University of Hannover, 30419 Hannover, Germany, [3]Centre for Plant Sciences, University of Leeds, LS2 9JT Leeds, UK and [4]Department of Agronomy, University of Rostock, 18059 Rostock, Germany

ABSTRACT

Sorghum is a promising alternative to maize for bioenergy production in Europe; however, its use is currently limited by poor adaptation to low temperatures during and after germination. We collected multi-trait phenotype data under optimal and suboptimal temperatures in a genetically diverse recombinant inbred line (RIL) mapping population showing contrasting segregation patterns for pre- and post-emergence chilling tolerance. Germination, emergence, seedling development, root architecture and seedling survival were assessed in two different seedlots. Emergence and root establishment were found to be the key determinants of development and survival under chilling stress. Highly interactive epistatic quantitative trait loci (QTL) hotspots, including a previously unknown QTL on Sb06 with a significant effect on prolonged chilling survival, were found to regulate different physiological mechanisms contributing to maintenance of growth and development despite the chilling temperatures. The major QTL regions harbour promising candidate genes with known roles in abiotic stress tolerance. Identification of loci in the QTL hotspot regions conferring maintenance of cell division and growth under early chilling stress represents a promising step towards breeding for successful establishment of sorghum in temperate climates.

Key-words: abiotic stress; chilling; low temperature; QTL; sweet sorghum.

INTRODUCTION

Sorghum (*Sorghum bicolor* L. Moench) is a C4 grass native to tropical and subtropical environments. C4 grasses are considered to be sensitive to cold temperatures; however, there is a growing body of evidence that C4 adaptation to warm, dry climates is attributable more to habitat changes, from shaded forest environments to drier woody or grassland areas, rather than to a *per se* preference for warmer environments (Edwards & Smith 2010). The adaptation of sorghum to drier environments led to its domestication in north-eastern Africa and subsequent establishment as the world's fifth most important grain crop. In the face of global climate change, reduced water and a need for improved nutrient use efficiency, sorghum is today playing an increasing role in meeting demand for animal feed and renewable energy biomass in many parts of the world.

Besides its high biomass potential and superior drought tolerance, the high stem sugar content, diabrotica resistance (Oyediran, Hibbard & Clark 2004) and nutrient use efficiency of sweet sorghum (Subbarao *et al*. 2006, 2009) make it a particularly interesting alternative to maize for bioenergy production (Reddy *et al*. 2008). As was the experience with maize, however, expansion of sorghum production into temperate climatic regions of northern Europe, America and northern Asia can only be achieved with adequate early-stage chilling tolerance, which can be defined as an ability of the plant to survive suboptimal temperatures during germination, emergence and establishment.

Because sorghum is a tropical plant, it does not tolerate frost. Chilling stress in this species generally sets in at temperatures below 20 °C, affecting several early developmental processes including emergence, seedling vigour and general metabolism (Peacock 1982; Chinnusamy, Zhu & Zhu, 2007). Fortunately *S. bicolor* contains considerable genetic variability for chilling stress sensitivity and minimum basal temperatures. Detailed phenotypic characterization and genetic dissection of these traits, accompanied by identification and recombination of useful variation, represent essential steps towards targeted introgression of suitable early-stage seedling establishment traits into high-yielding biomass forms for sorghum cultivation in temperate climates.

Previous genetic mapping studies of emergence and early-stage seedling development in sorghum (Knoll, Gunaratna & Ejeta 2008; Burow *et al*. 2011) were conducted under uncontrolled field conditions. Although these studies give valuable first insights into chromosome regions with potential roles in chilling adaptation, studying emergence and early stage seedling establishment under chilling in field conditions leads to strong and unpredictable quantitative trait loci (QTL)-by-environment interactions that complicate interpretation of the responses (Burow *et al*. 2011).

In the present study we performed extensive phenotyping for germination, emergence, seedling development, root architecture and seedling survival of sorghum seedlings under low-temperature conditions. The study was performed in a genetically diverse recombinant inbred line (RIL) mapping population that segregates strongly for pre- and

Correspondence: R. Snowdon. Fax: +496419937429; e-mail: rod.snowdon@agrar.uni-giessen.de

post-emergence seedling development under chilling conditions. Validation was performed by deep phenotyping of selected extreme genotypes, including evaluation of field emergence at two sowing times, to confirm the most important major effect and interaction QTL for key traits. The aim was to improve our understanding of the morphophysiological reactions and genetic basis of sorghum seedling development under chilling stress conditions. In the process, we identified interesting traits indicative of low-temperature injury, along with important genome regions controlling several developmental processes with high relevance for breeding.

MATERIALS AND METHODS

Plant materials

The study was performed in an *S. bicolor* RIL population derived from a cross between the sweet sorghum SS79 and the grain sorghum M71, described previously by Shiringani, Frisch & Friedt (2010). From a total of 213 F_7 RILs, a subset of 178 with fully mature and high-quality seed lots produced in Gross Gerau, Germany were used for phenotypic characterization in 2009, while the 2010 phenotyping used seeds from 200 F_7 RILs produced in Italy. Using seed lots derived from very different environments, we aimed to identify genetic components that contribute to development and survival under chilling stress independently of maternal effects on seed quality.

For the two seed lots a large series of contrasting experiments under controlled, semi-controlled and field conditions were performed in 2009 and 2010, respectively, to evaluate the genetic component of chilling stress responses over different treatments and seed lots. A total of 167 genotypes, plus the two parents, were represented in both seed lots. Subsequently, using seeds also produced in Gross Gerau, a chilling survival QTL validation experiment was conducted on 100 selected F_8 RILs that showed extreme responses to chilling in the previous experiments.

Overview of the experimental design

Chilling stress responses in the parental lines and mapping population were investigated in diverse phenotyping experiments assaying three important stages and processes (1) germination and emergence; (2) seedling root and shoot development; and (3) chlorophyll content before and after chilling stress. One of the most typical stress phenotypes observed under early chilling stress is chlorosis (often leading to death). We therefore considered the degree of reduction in chlorophyll content, whether through chlorophyll degradation or through reduced biogenesis in new leaves, to be a potential indicator of stress response and survival ability under low temperatures.

The different conditions used in the various experiments were designed to emulate potential scenarios for early spring temperatures in temperate climatic zones. Optimal (control) and chilling conditions were chosen according to previously reported basal temperatures for sorghum growth and development, considered to be 8.7 °C for germination and 14.5 °C for emergence and growth in a genotype-dependent manner (Brar *et al*. 1992; Payne, Batata & Rosenow 2003; Patanè *et al*. 2006; Fiedler *et al*. 2012). Full details of all experimental conditions and parameters used for the phenotyping are given as Electronic Supplementary Materials (Supporting Information Supplementary Methods). An outline of the exact stress and control conditions including the seed lot used for each experiment is provided in Table 1.

Unless specifically mentioned, all experiments were performed using a randomized complete block design with a minimum of two and a maximum of nine replications per genotype. The seeds for all soil-based assays and germination experiments were treated with Goucho® (Bayer Crop Science, Leverkusen, Germany) and Maxim® (Syngenta Crop Protection, Greensboro, NC, USA). Seed colour and thousand seed weight (TSW) were determined with a MARVIN seed imaging system (GTA Sensorik GmbH, Neubrandenburg, Germany).

Statistical analysis

One-way analysis of variance (ANOVA) and Pearson's correlation analysis were undertaken using IBM SPSS Statistics Version 18 (IBM Software, Armonk, NY, USA). A correlation heat map was generated with the software Hierarchical Cluster Explorer 3 (HCE3; Seo & Shneiderman 2002). Mixed models were calculated using the software SAS 9.1 (SAS, Cary, NC, USA) with genotypes set as random factors. Variance estimates were used to calculate broad-sense heritability (h^2) as follows:

$$h^2(\%) = [\sigma^2_G/(\sigma^2_G + \sigma^2/r)] \times 100$$

where σ^2_G is the genotypic variance, σ^2 is the error variance, and r is the number of replications.

A non-linear regression model was employed using the PROC non-linear regression (NLIN) procedure of SAS. This was done by setting bounds based on the population mean onset and cessation of emergence, along with cumulative emergence, allowing a model to be developed, which can estimate subsequent values. The convergence of the model was checked for individual genotypes, with adjustments performed to converge all entries. The following formulas were then used to estimate emergence-related traits:

$$S = \text{rate of emergence per day} = A_{max}/(T2 - T1)$$

$$T_{50} = \text{days until 50\% emergence} = [(T1) + (0.5 \times A_{max})/S],$$
$$\text{and Duration of emergence} = T2 - T1;$$

where
A_{max} = maximum cumulative emergence;
T1 = days until onset of emergence;
and T2 = days until cessation of emergence.

Table 1. Summary of conducted experiments, scored traits and set temperatures and light conditions

Year	Number of RILs	Experiment	Reps	Traits scored	Conditions	Temp °C	Day(h)/ night(h)	Length of experiment (days)
2009	178	Filter paper germination test	2	Days to onset of germination	Chilling	13	Constant dark	15
				Germination rate (%)	Optimum	30/20	8/16 dark	7
		Soil-based assay	3	Emergence rate	Chilling	13/11	11/13	30/60
				Root and shoot length Root and shoot biomass Survival rate (%)	Optimum	25/20	11 /13	9
			5	CCM	Optimum (before stress)	25/20	11/13	30
					1 week and 2 weeks after chilling stress	13/11	11/13	45
		Field experiment	2	Early sowing (end of April) Late sowing (end of May)	Soil temperature (see Supporting Information Fig. S1)			15–20
2010	198	Filter paper germination test	2	Germination rate (%)	Chilling	13	Constant dark	15
					Optimum	30/20	8/16 dark	7
		Gel-based root development	5	Root length	Chilling	13	Constant dark	15
					Optimum	30/20	8/16	7
		Soil-based assays	4–6	Daily emergence rate	CAS	13/10	12/12	40
				Root and shoot length Root and shoot biomass	CAE	Optimum for 5–6 DAS and 13/10 afterwards	12/12	40
				Survival rate (%)	Optimum	25 °C/20 °C	12/12	9
			5	CCM	Optimum (before stress)	25/20	11/13	30
					1 week and 2 weeks after chilling stress	13/10	11/13	45
2011	100	Sterile sand soil-based validation experiment	4	Survival (%) Root and shoot length Root and shoot biomass	CAE	Optimum for 4–7 DAS and 13/10 afterwards	12/12	40

CAE, chilling after emergence; CCM, chlorophyll content meter; DAS, days after sowing; RIL, recombinant inbred line.

Mapping and gene content of QTL associated with chilling stress reactions

QTL analysis was performed in the genetic map for SS79 × M71 developed by Shiringani *et al*. (2010), using the software PLABQTL (Utz & Melchinger 1996). QTL effects were reported from the final simultaneous fit of the regression analysis, following cofactor identification using stepwise regression and Akaki's selection criterion. The analysis mode, which considers additive × additive epistatic interactions was used for composite interval mapping. Main-effect QTL, epistatic QTL and their 1-LOD support intervals were displayed using the CIRCOS circular presentation software (Krzywinski *et al*. 2009).

The gene content within the support intervals of selected chilling stress-related QTL was assayed by alignment of the genetic map to the sorghum genome sequence (Paterson *et al*. 2009) via the *S. bicolor* annotation in the Phytozome database (http://www.phytozome.net/sorghum). Potential candidate genes within QTL support intervals were selected by gene enrichment analysis based on reported functions of annotated genes, using the software agriGO (Du *et al*. 2010). Peptide homologues of the most interesting candidate genes for key QTL were retrieved using the BioMart tool on Phytozome.

Organization, analysis and reporting of experiments in 2009, 2010 and 2011

Because of expected differences in seed vigour between the seed lots generated in Germany and Italy, the data from the 2 years of experiments using different seed lots were analysed separately. A more detailed analysis was performed for the 2010 experiments because of the superior characteristics of the seed lot, the larger population size ($n = 200$) and the large number of traits that were assessed in 2010. The population size directly influences the power of QTL detection and is therefore a particularly important consideration for understanding complex epistatic interactions between loci.

The first set of experiments in 2009 provided initial information for improvement of experimental designs, data acquisition and analysis strategies in the subsequent experiments. Hence, the results on the first experiment are only discussed in cases where we observed discrepancies, or if they had strategic importance for the latter, more substantial experiments. The third set of experiments, in 2011, was conducted to validate interesting patterns of prolonged chilling survival. The 2011 validation experiment was conducted under relatively sterile, but chilling stress conditions on selected genotypes ($n = 100$) that, during the 2010 experiments, showed phenotypes at the tail or head of the distribution for survival under prolonged chilling stress. Timing and developmental stages considered for most of the experiments were selected to simulate early stage seedling development of sorghum under chilling environments. The three important stages and processes studied were: germination and emergence, seedling root and shoot development and chlorophyll content upon chilling stress. Genetic analysis on the different traits from different batches under stress and optimum conditions were performed separately to identify genomic regions which are relevant across different seed batches, growth conditions and developmental processes. Overlapping QTL across years, traits and temperature conditions are considered useful for future marker assisted selection and improvement of early stage seedling vigour under temperate early spring conditions.

RESULTS

Trait variation and interrelationships

All the traits investigated showed significant differences ($P \leq 0.05$) under optimum and chilling conditions either between the parental accessions or within the RIL population (transgressive segregation). Details on the ranges of variation within each trait and heritability estimates under control and stress conditions will be presented in three categories: germination and emergence, seedling development and survival, and root architecture. To demonstrate the interrelationships among the various traits a cluster analysis based on Euclidean distances between traits in the 2010 trial is shown in Fig. 1. As expected, germination under both optimal and cold conditions clustered with biomass traits under optimal conditions. On the other hand, emergence was found to cluster with survival, biomass and chlorophyll concentrations under both optimum and stress conditions. Correlation analyses among all traits and treatments (Supporting Information Table S1a) revealed that two of the most agronomical important traits, emergence and survival under cold temperature, showed significant correlations with numerous other traits (Fig. 2). The survival rate under cold sowing and growth conditions was correlated with all traits except chlorophyll content meter (CCM) values, root dry weight (DW) and germination. In contrast, survival under exposure to chilling after emergence (CAE) was correlated with CCM values after 2 weeks of chilling stress, but with few other traits. Emergence under chilling stress showed interesting correlations with germination, root length, and shoot fresh weight (FW) and DW, respectively. As expected a positive correlation was observed between cold-temperature emergence and the survival rate under exposure to CAE. On the other hand no significant correlation was found between cold emergence and CAE survival. The correlation of cold emergence rate to survival was higher ($r = 0.601$) than it was to cold germination ($r = 0.40$).

The CCM values were found to be correlated with CAE survival rate, the root to shoot ratio, root biomass and root length. The rate of seedling emergence (S) correlated with several other traits, reflecting a common influence of cell division and elongation. Root length and biomass at optimum temperature were correlated with both chilling at sowing (CAS) and CAE survival. The correlation between CAS root length and CAS survival was negative, reflecting the greater influence of root biomass (number of lateral roots) on cold survival (as seen in the parental lines) rather than on primary root length. Whereas survival and root biomass were correlated under both cold and optimum

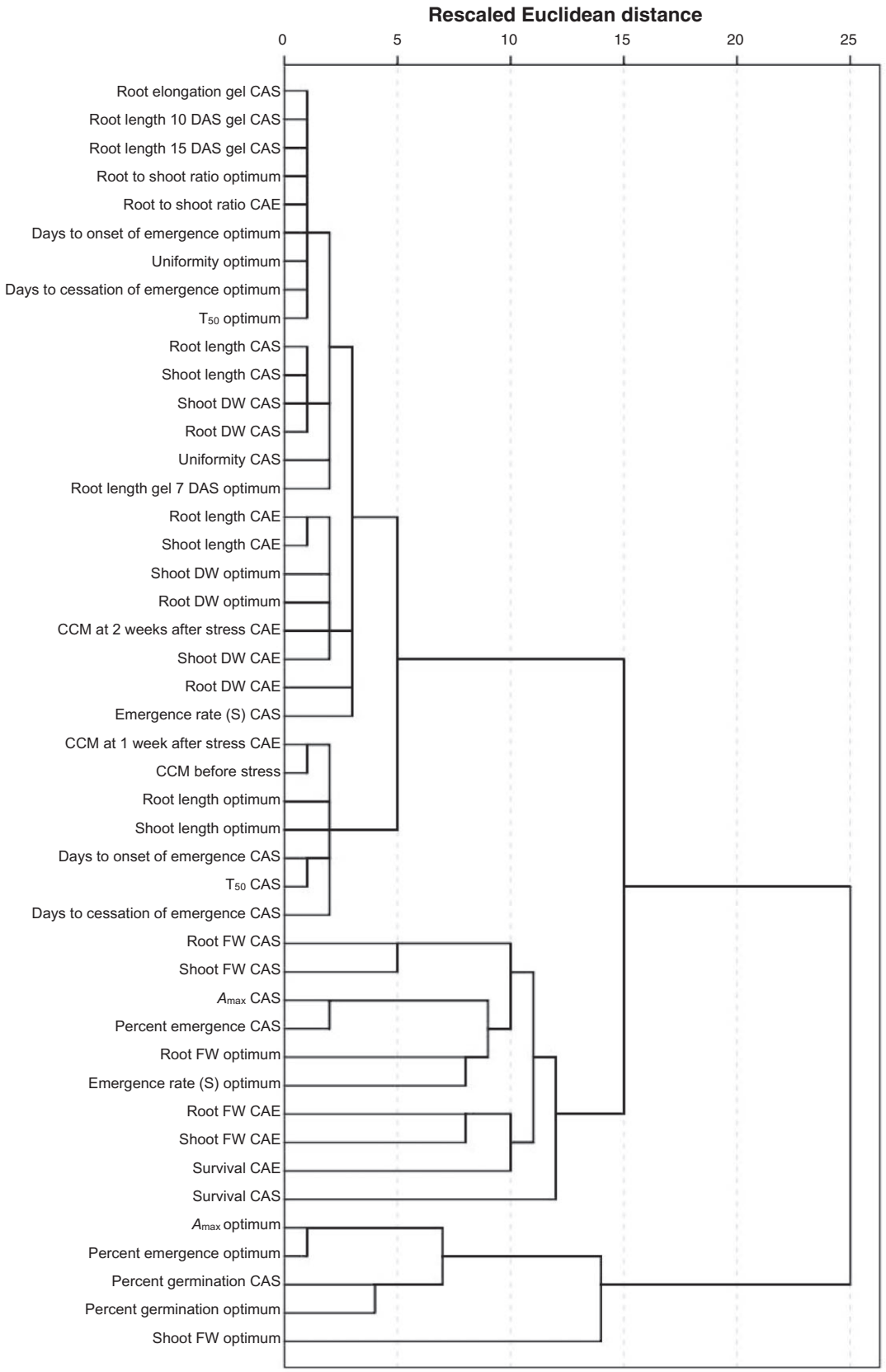


Figure 1. Euclidean distance-based cluster analysis of traits assessed in the 2010 greenhouse evaluations of the SS79 × M71 recombinant inbred line (RIL) population ($n = 200$) under optimal and suboptimal temperature conditions. CAE, chilling after emergence; CAS, chilling at sowing; CCM, chlorophyll content meter; DAS, days after sowing; DW, dry weight; FW, fresh weight.

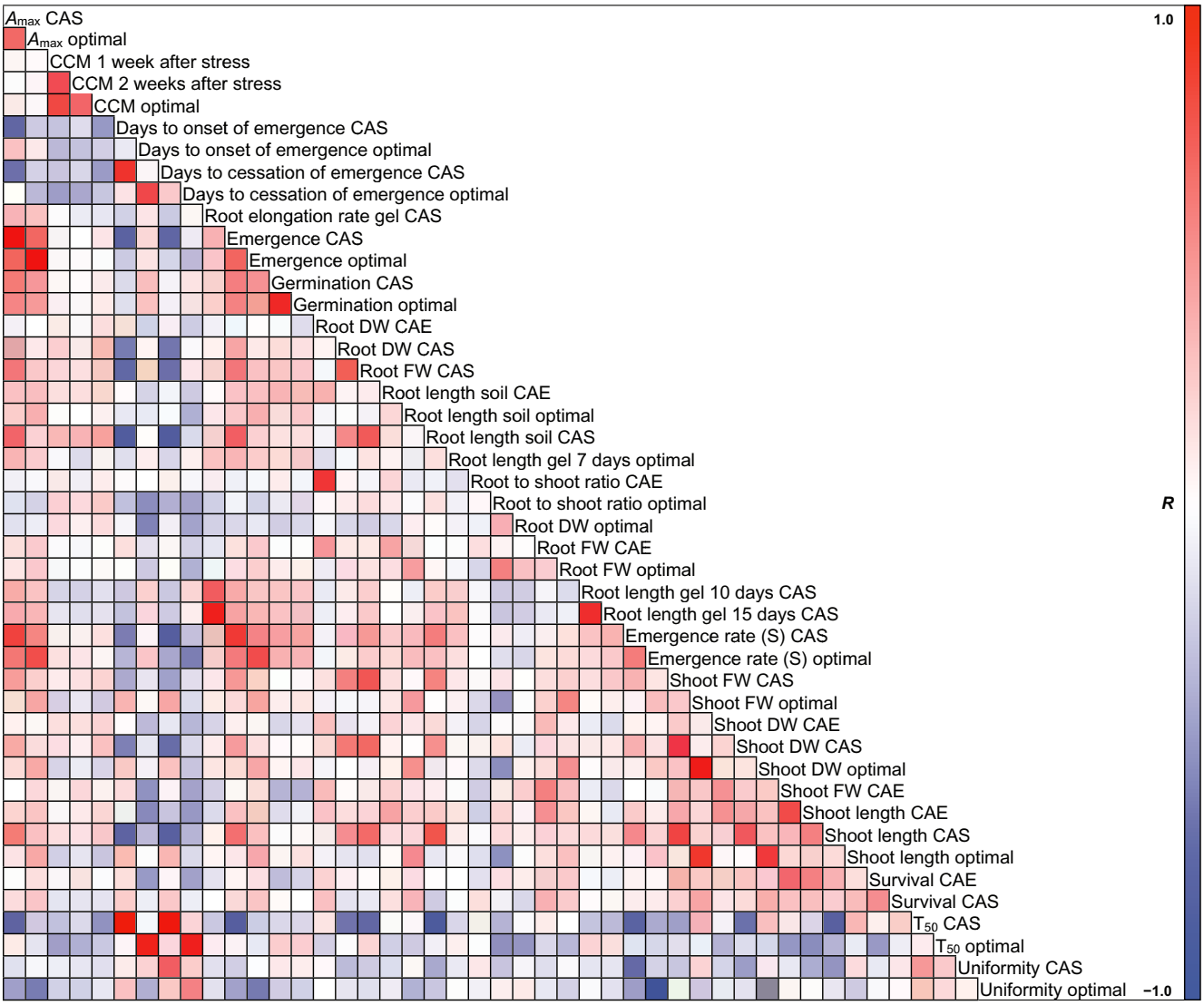


Figure 2. Heat map showing Pearson's correlations (r) among all traits assessed in the 2010 greenhouse evaluations of the SS79 × M71 recombinant inbred line (RIL) population ($n = 200$) under optimal and suboptimal temperature conditions. Full details of trait correlations from all experiments are given in Supporting Information (Table S1). CAE, chilling after emergence; CAS, chilling at sowing; CCM, chlorophyll content meter; DW, dry weight; FW, fresh weight.

conditions, the shoot biomass was not correlated to survival under cold conditions.

Similar correlation analysis on the 2009 data showed controlled environment and field germination and emergence results are correlated and root biomass was correlated with field emergence in locations where the soil temperature was lower (Supporting Information Table S1b). Linear regression analysis using cold emergence in soil (2009 experiments) as a dependent variable and TSW, seed colour, days to onset of germination and cold germination as emergence-related independent variables, gave a statistically significant model ($P < 0.01$) with $r^2 = 0.47$. The two independent factors seed colour and TSW were not statistically significant ($P < 0.05$). Cold germination and days to onset of germination under chilling stress had standardized beta values of 0.17 and 0.76, respectively. Together, the multiple linear regressions indicated the low influence of seed characteristics on emergence in soil in these materials.

Effect of chilling stress on germination and emergence

Mean germination and emergence rates of SS79, M71 and the RIL population on soil media under chilling and optimum conditions are indicated in Table 2. ANOVA revealed statistically significant differences ($P \leq 0.01$) in germination and emergence rates caused by temperature, genotype and genotype-temperature interactions, respectively. Combined analysis of the two seed lots, which showed significant differences ($P < 0.01$) in TSW and seed colour, also indicated

Table 2. Summary of germination and emergence data for chilling tolerance experiments in SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population, using two seed lots from 2009 and 2010, respectively

		Parental means		RIL population				
Experiment	Trait	SS79	M71	Replications	Mean	Min.	Max.	h^2 (%)
2009 (178 RILs)	Germination (%) optimal	67.00	78.00	2	72.94	3.00	99.00	60.0
	Germination (%) cold	43.00	70.02	2	58.63	0.00	99.00	88.1
	Emergence optimal 9DAS	58.33	80.56	3	66.94	5.56	100.00	82.6
	Emergence cold 1MAS	50.00	38.89	9	49.97	0.00	97.22	87.1
	Seed colour (field trial)	4.35	1.93	2	3.61	2.16	6.04	88.0
	TSW (g, field trial)	14.47	15.23	2	19.16	4.41	30.25	94.0
2010 (198 RILs)	Germination optimal (%)	91.68	92.00	2	83.08	0.00	100.00	93.7
	Germination cold (%)	81.00	96.00	2	78.67	0.00	100.00	96.5
	Emergence optimal	87.65	61.73	6	74.63	12.04	98.15	91.0
	Emergence cold	73.33	32.81	2	44.76	0.00	92.19	92.0
	Seed colour (field trial)	3.70	2.96	2	3.13	2.06	5.56	–
	TSW (g, field trial)	19.22	25.18	2	22.13	10.33	31.23	–
Mean over all replications	Germination (%) optimal	79.34	85.00	4	78.01	1.50	99.50	–
	Germination (%) cold	62.00	83.01	4	68.65	0.00	99.50	–
	Emergence optimal	72.99	71.14	9	70.78	8.80	99.07	–
	Emergence cold	61.67	35.85	11	47.37	0.00	94.71	–
	Seed colour (field trial)	4.02	2.45	4	3.37	2.11	5.80	–
	TSW (g, field trial)	16.84	20.20	4	20.65	7.37	30.74	–

DAS, days after sowing; MAS, months after sowing; *N*, number of RILs phenotyped; relative (%) performance of chilling in relation to optimum conditions; TSW, thousand-seed weight; h^2, broad-sense heritability.

significant interactions ($P < 0.01$) between genotype and seed lot. The grain sorghum parent M71 showed a significantly stronger sensitivity to emergence under chilling conditions compared with the sweet sorghum parent SS79. Under optimal temperature conditions SS79 had a slightly lower mean germination rate (79%) than M71 (85%) using the 2009 seed lot from Gross Gerau, whereas no difference was seen with the 2010 seed lot from Italy. In the 2009 experiment, the emergence of M71 under optimal conditions was higher than that of SS79; however, the cold emergence of M71 was considerably lower than SS79 ($P \leq 0.05$) for both seed lots (Table 2).

Table 3 summarizes the emergence rate experiment, describing the most important parameters analysed using the NLIN function based on emergence rate data collected from sowing to 29 days after sowing (DAS) under cold conditions. Ample variation was observed within the RIL population, with the parents segregating for different emergence rate parameters such as velocity of emergence (S) and maximum emergence (A_{max}) under low temperature. Velocity of emergence was one of the traits with the greatest variation within the population, while A_{max} also showed a very high standard variation of 23%. As can be seen from Table 3, the onset and cessation of emergence was delayed

Table 3. Results summary from the non-linear regression (NLIN) analysis of emergency (% of sown seeds) from the 2010 greenhouse experiment with SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population

		Parental means		RIL population (n = 196)			
Treatment	Parameter	SS79	M71	Mean	Min	Max	SD
Cold (2 replications)	T1 (onset of emergence, d)	14.42	15.43	15.77	11.97	26.00	2.85
	T2 (cessation of emergence, d)	20.00	19.94	22.07	15.31	27.00	4.69
	V [velocity of emergence: A_{max}/(T2 – T1)]	11.37	6.90	8.41	0.00	47.00	6.68
	A_{max} [maximum emergence: S × (T2 – T1)]	63.46	31.10	42.89	0.00	90.38	23.54
	Uniformity (T2 – T1)	5.58	4.51	5.87	1.00	12.93	2.05
	T_{50} (time to reach to 50% emergence, d)	17.21	17.69	18.73	14.10	26.62	3.11
Optimum (6 replications)	T1 (onset of emergence, d)	2.36	3.96	2.24	0.76	3.72	0.69
	V [velocity of emergence: A_{max}/(T2 – T1)]	30.5	54	46.16	2.60	89.00	17.43
	T2 (cessation of emergence, d)	5.21	5.10	4.01	2.03	6.23	0.76
	A_{max} [maximum emergence: S × (T2 – T1)]	87.00	61.67	73.56	12.00	97.75	14.04
	Uniformity (T2 – T1)	2.85	1.14	1.77	1.03	4.62	0.55
	T_{50} (time to reach to 50% emergence, d)	3.79	4.54	3.12	1.51	4.52	0.67

S, rate of emergence per day; SD, standard deviation.

Table 4. Summary of variation field emergence (% of sown seeds) after different sowing dates in trials of SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population, at three locations in Germany during the 2009 growing season

Location	Sowing date	Parental means		RIL population (n = 178)				h^2 (all locations)
		SS79	M71	Mean	Min	Max	SD	
Gross-Gerau	30.04.2009	82.50	55.00	59.21	0.00	100.00	32.11	68.89
	03.06.2009	67.50	30.00	59.13	5.00	100.00	30.96	
Giessen	20.05.2009	70.00	60.00	64.71	7.50	95.00	19.82	
	28.05.2009	15.00	50.00	23.97	0.00	80.00	17.00	
Hannover	30.04.2009	27.50	7.50	14.67	0.00	58.75	13.19	
	11.05.2009	43.75	6.25	23.62	0.00	90.00	19.29	

Two replications per line per sowing date were performed at each location. SD, standard deviation; h^2 = broad-sense heritability (%).

by chilling, resulting in longer duration of emergence under cold conditions.

The summary of field emergence (2009) in Table 4 shows that SS79 has better emergence than M71, except the second sowing time at the location Giessen. This was due to a drop in temperature after emergence of the late-sown trial (see Supporting Information Fig. S1). In general, the warmer mean soil temperatures at Gross Gerau resulted in higher emergence than at the other two locations.

Effects of chilling stress on seedling development and survival

The summary table for the seedling development related traits shows the transgressive segregation observed for most of the traits considered (Tables 5 and 6). The parental lines segregate for most of the traits including the survival of the seedlings under prolonged chilling conditions (Fig. 3). Root length and biomass are correlated; however, at the parental genotype level the longer root of SS79 is not translated as higher root biomass, rather M71 has shorter axial roots but large number of lateral roots that gave rise to a higher root biomass.

The leaf chlorophyll content (CCM value) of M71 was higher under optimal conditions than that of SS79, and a very high variation was observed within the RIL population (Tables 5 and 6). The reduction in chlorophyll content during the 2 weeks of chilling stress was considerably higher in M71 than SS79, and the RIL population showed considerable variation in the degree of chlorophyll degradation upon stress.

The survival rate after 40–60 days chilling stress was higher for M71 than for SS79 and the population mean (Table 7). The mean plant survival rate was lower when plants were stressed after warm emergence (CAE) than in the cold emergence experiment (CAS). The survival of M71 was significantly higher than that of SS79 under CAS ($P < 0.05$). High

Table 5. Summary statistics for the 2009 growth chamber chilling evaluation experiments with SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population

Treatment	Trait	Parental means		RIL population (n = 178)				h^2 (%)
		SS79	M71	Mean	Min	Max	SD	
Cold	Days to onset of germination	5.5	6.0	5.3	3.0	11.0	1.0	31.0
	Shoot length at 2 MAS (cm)	5.5	6.0	6.4	2.2	9.7	1.3	82.5
	Root length at 1 MAS (cm)	2.8	2.5	3.8	1.1	7.4	1.4	84.7
	Shoot FW at 1 MAS	28.3	51.5	43.4	0.0	267.0	22.7	53.6
	Shoot DW at 1 MAS	3.6	5.0	4.5	2.1	18.4	1.7	13.4
	Shoot FW at 2 MAS	16.4	62.0	38.8	0.0	7.7	14.1	54.6
	Shoot DW at 2 MAS	3.6	6.8	5.3	0.0	13.0	1.5	47.7
	Root FW at 1 MAS	10.5	17.5	23.7	0.0	94.2	10.8	29.7
	Root DW at 1 MAS	2.1	1.9	3.4	0.0	9.5	1.6	64.8
	Chlorophyll content (CCM) at 1 WAS	5.8	4.4	4.4	2.1	10.1	1.2	79.5
	Chlorophyll content (CCM) at 2 WAS	4.5	4.1	3.9	2.1	8.3	0.9	76.3
CAS	Shoot length (cm) at 1 MAS	4.6	3.6	4.1	2.1	6.8	1.0	89.7
Optimum	Days to onset of germination	1.5	1.5	1.4	1.0	3.0	0.4	61.0
	Shoot length (cm) at 9 DAS	13.7	17.0	12.8	6.2	19.4	2.6	92.1
	Shoot FW at 9 DAS	10.1	18.1	11.9	3.0	14.2	2.5	87.9
	Shoot DW at 9 DAS	9.3	13.5	10.6	0.0	54.0	6.6	66.6
	Chlorophyll content (CCM) before stress	7.4	7.6	6.3	3.2	12.4	1.7	78.1

DAS, days after sowing; MAS, months after sowing; WAS, weeks after stress; CAS, chilling after sowing; FW, fresh weight (mb/plant); DW, dry weight (mg/plant); SD, standard deviation; h^2 = broad-sense heritability.

Table 6. Summary statistics for the 2010 greenhouse chilling evaluation experiments with SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population

			Parental means		RIL population (n = 178)				
Treatment	Replications	Trait	SS79	M71	Mean	Min	Max	SD	h^2 (%)
Optimum	4	Shoot FW (mg/plant)	79.6	85.4	100.0	37.0	182.9	27.2	90.6
		Shoot DW (mg/plant)	8.2	8.0	9.8	3.7	17.3	2.5	87.0
		Root FW (mg/plant)	38.1	74.8	50.5	18.7	104.3	14.1	85.5
		Root DW (mg/plant)	11.8	8.0	10.1	3.6	22.9	3.2	84.2
		Shoot length (cm)	11.8	10.9	13.7	8.7	19.8	2.3	93.5
		Root length (cm)	13.5	14.7	15.9	8.5	20.8	1.7	71.3
CAS	2	Shoot FW (mg/plant)	32.0	38.9	32.9	4.4	86.4	13.2	59.1
		Shoot DW (mg/plant)	3.9	4.8	4.9	0.6	12.6	1.7	41.5
		Root FW (mg/plant)	24.3	41.8	25.3	4.0	69.3	12.5	70.6
		Root DW (mg/plant)	4.7	8.5	5.4	0.0	24.8	3.1	29.2
	6	Shoot length (cm)	6.8	5.0	5.7	1.7	10.5	1.8	84.7
		Root length (cm)	7.8	5.5	5.8	1.0	10.5	2.1	85.4
CAE	4	Shoot FW (mg/plant)	34.3	47.1	46.8	5.0	122.3	19.5	66.2
		Shoot DW (mg/plant)	7.4	7.0	8.6	2.2	39.6	4.1	52.0
		Root FW (mg/plant)	20.1	48.7	36.9	3.0	196.0	20.0	69.0
		Root DW (mg/plant)	6.1	9.6	8.5	2.6	58.7	6.0	73.7
		Root length (cm)	10.0	10.8	10.2	6.0	22.9	2.1	21.5
		Shoot length (cm)	9.4	8.2	8.7	1.0	14.1	2.1	72.0
	5	CCM before chilling stress	17.2	17.0	16.7	9.2	24.6	2.8	81.9
		CCM 1 WAS	12.2	13.2	14.4	6.6	22.1	3.0	86.7
		CCM 2 WAS	11.3	10.0	11.9	4.1	24.0	3.5	82.3

FW, fresh weight; DW, dry weight; CAS, chilling at sowing; WAS, weeks after stress; CAE, chilling at emergence; SD, standard deviation; h^2, broad-sense heritability.

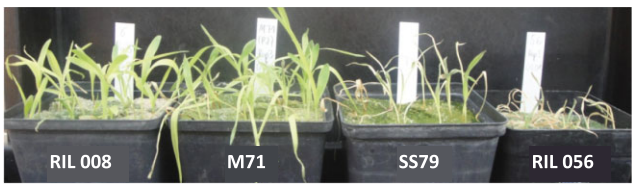


Figure 3. Survival of seedlings from parental lines and extreme-phenotype recombinant inbred lines (RILs) after growth under chilling conditions (13 °C/10 °C) for 35 days after emergence during the CAE validation experiment. SS79 and M71 are the two parental lines, while RIL 56 and RIL 08 show transgressive segregation for survival under prolonged chilling stress.

variability in survival rate was observed within the RIL population, with values ranging from 0 to 93% in both the CAE and CAS experiments. CAE survival showed a higher heritability compared with the CAS, although the lower calculated heritability for CAS may be due to this experiment being replicated only twice. Taking the two sowing conditions as a factor in the GLM ANOVA indicated that the sowing temperature, as expected, significantly influences survival at $P < 0.01$. On the other hand, our data indicate a great degree of genetically determined variation for response to low temperature within our RIL population that is of great interest for breeding towards improved survival under early-stage chilling stress.

Table 7. Summary of seedling survival rates in three independent experiments conducted with SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population under controlled conditions from 2009 to 2011

			Parental means		RIL population						
Year	Treatment	Stress time point	SS79	M71	N	Reps	Mean	Min	Max	SD	h^2 (%)
2009	CAS	Starting from sowing	41.18	76.47	178	–	46.9	0.0	100.0	39.2	–
2010	CAS	Starting from sowing	23.33	86.37	196	2	42.1	0.0	100.0	30.4	81.1
	CAE	Immediately after emergence	38.37	46.30	196	4	25.6	0.0	92.6	26.0	92.0
2011	CAE	Chilling after 4–5 days	23.72	47.25	95	2	18.2	0.0	80.0	20.0	79.2
		Chilling after 6–7 days	58.34	62.11	95	2	38.7	0.0	96.7	23.3	69.9
		Validation mean (3–7 days)	41.02	54.68	95	4	28.8	15.3	72.2	19.3	72.2

Variance analysis and broad sense heritability (h^2) could not be calculated in the 2009 and 2011 experiments because survival was scored together for all nine minipots. CAS, chilling at sowing; CAE, chilling after emergence; DAS, days after sowing; Reps, number of repetitions; *n*, number of RILs tested; SD, standard deviation; h^2, broad-sense heritability.

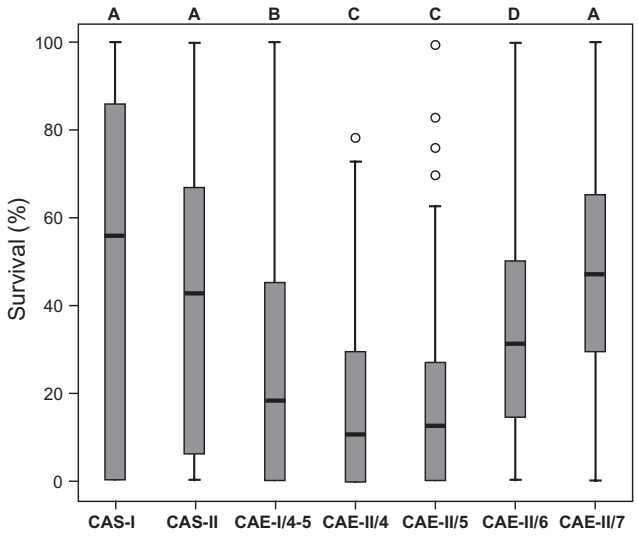


Figure 4. Differences in mean survival of SS79 × M71 recombinant inbred lines (RILs) in independent experiments after exposure to chilling stress (13 °C/10 °C day/night) at zero days CAS compared with chilling after emergence beginning 4–5 days after emergence (CAE-I/4–5) or between 4 and 7 days after emergence (CAE-II/4 to CAE-II/7), respectively. Different letters above the distribution bars represent significant mean square differences ($P \leq 0.05$) in survival between the different treatments. CAE, chilling after emergence; CAS, chilling at sowing.

Similar parental survival mean values, segregation and population distribution were observed in the 2009 experiments and the 2011 chilling survival validation experiment (Fig. 4). The survival data confirmed the very strong correlation between several traits, including root fresh and dry matter yield (Supporting Information Table S1c). Interestingly, combined ANOVA and mean separation revealed that chilling at 4–5 days after emergence (2010 and 2011) resulted in significantly reduced survival compared with chilling at emergence or after 6–7 days (Fig. 4).

Root architecture and development under chilling conditions

At both 10 and 15 DAS, SS79 showed a considerably longer mean root length than M71. Under optimal conditions the two parental lines showed no statistically significant difference in seminal root length, although ample variability was observed within the RIL population both under chilling stress and optimal temperatures (Table 8). Examination of heterotrophic and autotrophic growth after 1 week of recovery under warm conditions showed that M71 developed a large number of lateral and seminal roots both during the stress application and during the recovery phase (Supporting Information Fig. S2).

QTL analysis of seedling germination and growth under chilling stress

As in the earlier, the QTL results from the 2010 experiments are presented in detail and constitutive QTL detected in comparison with 2009 and the 2011 selective phenoyping data will be presented.

In the greenhouse experiments in 2010, 68 main-effect QTL and 101 epistatic QTL were identified. Complex epistatic interactions were observed throughout the whole genome and significant hotspots of multi-trait epistasis were detected (Fig. 5). Details of all main-effect and epistatic QTL detected for all experiments are provided in Supporting Information Tables S2, S3 and S4, respectively.

The main-effect and epistatic QTL for survival in the CAS experiment explained 92.8% of the total variance for this trait. One major main-effect QTL on chromosome Sb07 had 21.8% effect on chilling survival in the CAS experiment. Even though it has only one main effect QTL, cold germination has the second highest total variance explained by the QTL. In contrast, the total variance explained by all nine QTL detected for root length in the CAE experiment explained only 40% of the total phenotypic variance. This lack of phenotypic effects from the main-effect QTL is explained by the large number of epistatic QTL for cold germination and the complete absence of epistatic QTL for root length in the CAE experiment.

A total of 101 digenic espistatic QTL were detected in 30 traits of the 2010 analysis. Individual interactions explained between 2.8 and 24% of the total phenotypic variance. The maximum number of epistatic interaction QTL detected per trait was 13 in case of germination under cold conditions followed by survival in the CAS experiment and cold emergence.

Table 8. Summary statistics for root development (primary root length in cm, 5 replications) in SS79, M71 and the SS79 × M71 recombinant inbred line (RIL) population measured in the gel-based assay under optimum and cold conditions in 2010

		Parental means		RIL population (n = 190)				
Treatment	Trait	SS79	M71	Mean	Min	Max	SD	h^2 (%)
Optimum	Root length 7 DAS	5.47	6.09	6.79	0.34	14.57	3.49	69.51
Cold	Root length 10 DAS	1.83	0.54	0.46	0.12	1.46	0.24	74.30
	Root length 15 DAS	3.05	0.88	0.88	0.14	3.80	0.58	71.16
	Root elongation rate (Δ root length)	1.22	0.34	0.42	0.10	2.65	0.39	–

DAS, days after sowing; SD, standard deviation; h^2, broad-sense heritability.

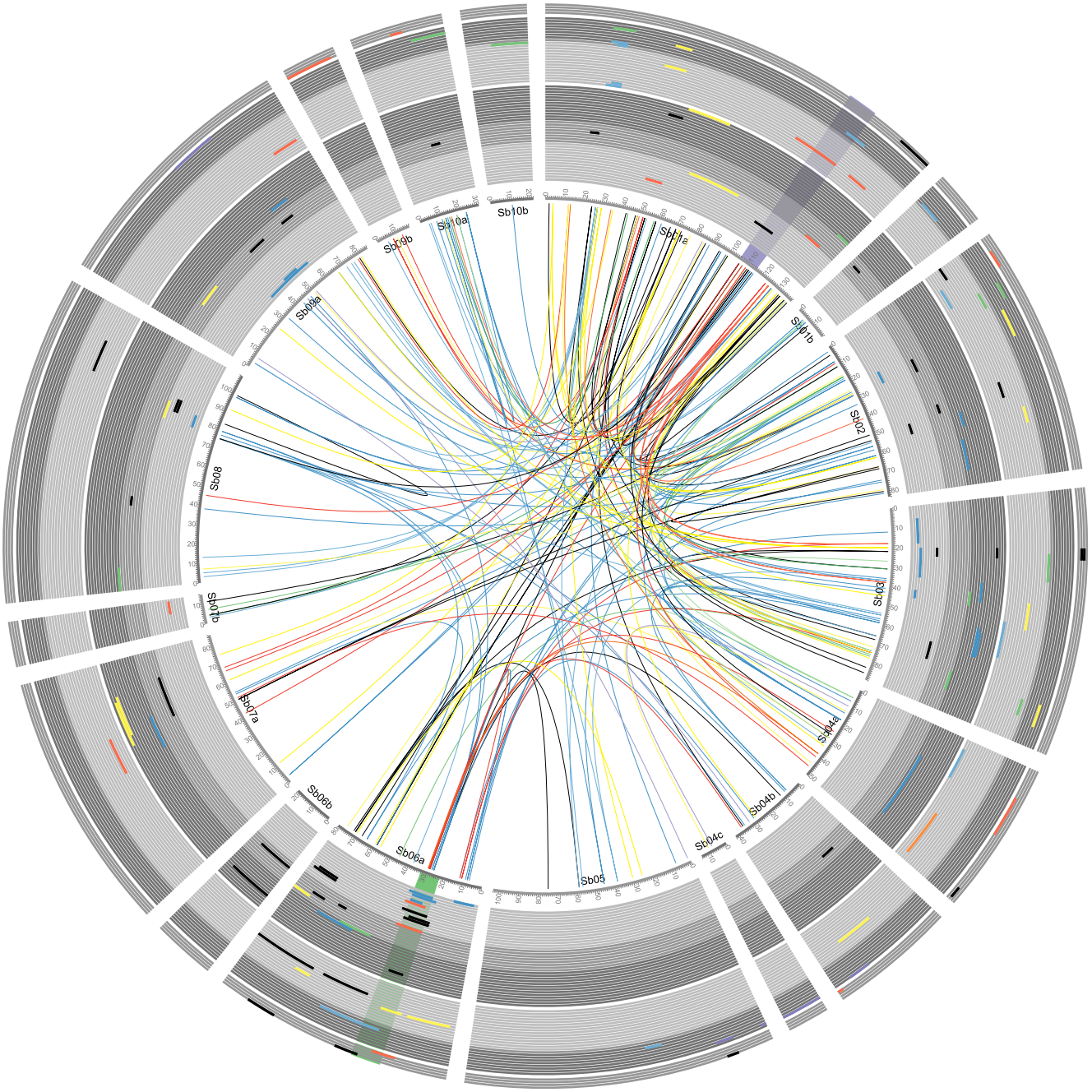


Figure 5. Circular genetic map from the *Sorghum bicolor* cross SS79 × M71 showing positions of main-effect QTL and two-locus epistatic QTL (internal connecting lines) for all greenhouse, climate chamber and field experiments conducted in 2010, 2009 and 2011. The light grey tracks represent traits measured after chilling at sowing (CAS), medium grey tracks represent traits measured during chilling after emergence (CAE) experiments, and dark grey tracks represent traits measured under optimum conditions and the white tracks show year boundaries. The coloured bars show the support intervals of QTL. Blue bars and connecting lines represent QTL and epistatic interactions for germination and emergence related traits; black: root traits; red: survival; yellow: shoot traits; green: leaf chlorophyll traits. A full list of the specific traits numbered from 1 (innermost track) to 63 (outermost track) is provided in Supporting Information (Fig. S3). The highlighted genome parts were the regions showing the highest degree of colocalization for main effect QTL (Sb06 at 30cM) and epistatic QTL (Sb01 at 110cM), respectively.

As expected from segregation pattern of the parents, QTL alleles with positive effects were contributed by SS79 for rate of emergence (S), root length under cold stress, elongation rate and rate of emergence. Alleles conferring improvements in other traits, including survival, were contributed by both parents; these are visible in Supporting Information (Tables S2a and S3a) as positive and negative additive effects for M71 and SS79, respectively.

Interesting QTL detected across experiments, traits and temperatures

Of all the traits considered, emergence related traits, root development traits and survival were found to co-locate in several chromosomes (Sb01, Sb02, Sb03, Sb06, Sb08 and Sb10). This is in agreement with our expectations and the correlation observed between these traits. Despite differences in the number of genotypes and seed production location that resulted in different seed vigour, interesting consistent QTL hubs were detected on several chromosomes. As is clearly shown in Fig. 5, there are at least two epistatic QTL hubs on the top and bottom of chromosome Sb01. The bottom QTL hub, close to the main effect QTL for root biomass (110cM), is an epistatic QTL for several emergence and root development traits. This same location was detected in 2009 as an epistatic QTL for emergence and a main effect QTL interval for survival and emergence, respectively. This region was again detected during the validation experiment as an epistatic QTL hub, explaining 37.1% (r^2) of the phenotypic variation for survival and interacting with three other QTL on chromosome Sb01 and Sb03. An epistatic QTL explaining 18% of the phenotypic variance for root FW was also detected in the same region. The prolonged survival validation experiment identified one main effect QTL explaining 27% of the phenotypic variance, on chromosome Sb02 (5-10cM). This region was also detected during the 2009 trial as a main effect QTL for onset of germination under chilling conditions and for chlorophyll content after 1 week of stress; it also appears as an epistatic QTL for root length and dry matter under chilling. In 2010, the same region harboured QTL for root to shoot ratio under chilling conditions, and epistatic QTL for percent germination, onset of emergence under optimum conditions and the time required to reach to 50% emergence under optimal conditions. Another interesting multi-trait QTL hub is on chromosome Sb03 around 20cM, where main effect and epistatic QTL for emergence and root development traits coincide. The same region was also detected in 2009 as a main effect QTL for chlorophyll content and as an epistatic QTL for seedling height. During the validation experiment, this region was an epistatic QTL explaining 21% of the phenotypic variance for survival and interacting with the above mentioned epistatic loci on chromosome Sb01. It is also a main effect QTL for both root fresh and DW. This location is close to the stay green locus Stg2 (Harris *et al*. 2007). Furthermore, on chromosome Sb06 at around 20–30 cM another interesting region harbours QTL for emergence, root development and survival under chilling stress conditions (Fig. 5). In 2009, this same region was one of three epistatic QTL for field emergence and days to onset of germination under optimum temperature. A main effect QTL for shoot length under chilling conditions (seedling vigour) was also mapped close to this region. The survival CAE validation experiment using selected lines showed that this region is a QTL hub containing overlapping main effect QTL influencing survival, root and shoot length.

Important stress tolerance candidate genes are located in major QTL hotspots

Alignment of markers flanking the most important QTL to the *S. bicolor* reference genome sequence enabled identification of a number of extremely promising candidate genes, with known roles in abiotic stress responses in maize and other plants, within the main-effect and epistatic QTL for chilling tolerance. A list of candidate genes with a potential involvement in abiotic stress tolerance, found in the major QTL hotspots on chromosomes Sb01, Sb04 and Sb06, is given in the Supporting Information (Table S5).

DISCUSSION

Genetic potential of sorghum to adapt to early spring temperate conditions

Early-stage chilling tolerance is a prerequisite for the expansion of sorghum in the northern hemisphere, especially if we expect sorghum to be competitive with adapted biomass crops like maize as an energy feedstock. Several methods have been used to determine early stage chilling response in sorghum and other tropical crop plants such as maize (Brar *et al*. 1992; Lee *et al*. 2002; Payne *et al*. 2003; Yu *et al*. 2003; Hund *et al*. 2004; Zhi-Hong *et al*. 2005; Patanè *et al*. 2006; Fiedler *et al*. 2012). In this study, we performed extensive phenotyping using many different approaches, in order to gain an understanding of the multi-trait interactions influenced by chilling stress and to identify traits that facilitate higher genetic gains per selection in breeding for chilling tolerance. The high heritability and the genetic and phenotypic correlations among several traits, particularly chilling survival, emergence and root development traits, indicate a strong potential for breeding using surrogate traits for rapid selection gains.

Extensive genetic and phenotypic variability, confirmed in our phenotyping study, enables sorghum to adapt to a range of altitudinal and latitudinal clines ranging from eastern Africa to Southeast Asia, Australia and the Americas (Brar *et al*. 1992; Yu *et al*. 2003; Knoll *et al*. 2008; Burow *et al*. 2011). The highlands of Africa, including the origin of our mapping population in southern Africa, are considered one of the most important sources of sorghum germplasm with adaptation to cool temperatures (Balota *et al*. 2010).

Through extensive phenotyping of a large genetic mapping population, we have clearly shown that seedling development processes in sorghum interact strongly with the environment. The parental genotypes used in our study showed interesting segregation patterns under low temperature conditions, both during and after emergence. Root establishment was found to be one of the most important factors influencing field establishment and survival under prolonged chilling stress conditions, regardless of the germination or emergence conditions.

Low or fluctuating temperatures pose major problems for field establishment. In agreement with other reports on emergence rate in sorghum (e.g. Fiedler *et al*. 2012), our seedling emergence NLIN models showed that chilling stress delays

the onset and speed of emergence from soil, as well as increasing the duration of emergence. Our results also demonstrated that filter paper-based germination assays are not always effective indicators for germination and emergence under chilling conditions in the field, or even in sterile soil. Filter paper germination of M71 was found to be higher than SS79 under both cool and optimal temperature conditions; however, the situation was reversed during soil-based assays under chilling conditions. The lower cold emergence rate of M71 was confirmed even under sterile soil conditions. Seed lot variation was shown to influence seed vigour, and the parental lines segregate for seed characteristics such as seed colour and size; however, seed colour differences were not associated with the differences in emergence at a population level. Obtaining a better understanding of the genetics and physiology of early stage developmental processes is one interesting area of future research into which our results give first insights. Because the two seed lots were produced in Germany and Italy with high mean temperature differences during seed production, the difference in seed vigour is attributed to the effect of the maternal environment on the physiological state of the seed at harvest. We confirmed previous findings showing correlations of emergence to several traits determining plant size as well as survival under chilling conditions (e.g. Balota *et al*. 2010; Mercer, Alexander & Snow 2011). This is reflected by co-localization of QTL for cold emergence, survival, emergence rate and root/shoot length, and biomass. These findings suggest possible pleiotropic action of responsible genes, or linkage between multiple genes controlling interrelated traits under stress and optimal conditions. Corresponding co-localization of QTL for emergence and seedling vigour has been reported in sorghum and maize seedlings grown under field conditions (Knoll *et al*. 2008; Burow *et al*. 2011). Interestingly, the parental line M71 showed lower emergence, but higher survival rate than the other parent, SS79. In contrast, positive correlations were observed in the RIL population between emergence and many biomass and survival traits, as described by Menkir & Larter (1987) and Yu *et al*. (2003) for maize and sorghum in controlled-environment and field experiments.

Despite the variable number of genotypes, slightly different experimental set ups and seed batches from contrasting environments, a number of constitutive QTL were detected across different years, conditions and traits. Interesting QTL hubs on chromosomes Sb01, Sb02, Sb03, Sb04 and Sb06 (Supporting Information Fig. S3) are key starting points for future fine mapping and candidate gene validation experiments, and for future application of marker-assisted selection for early-stage seedling development improvement.

Emergence and seedling establishment are important to adapt sorghum to temperate climates

Tropical annual crop plants such as sunflower, rice, maize, sorghum and tomato are generally chilling sensitive, especially at their emergence and early seedling developmental stages (Andaya & Mackill 2003; Hund *et al*. 2004; Knoll *et al*. 2008; Mercer *et al*. 2011). Germination and emergence rates are important under natural conditions for synchronization of emergence with environmental cues and competing species, and for determination of plant size (Mercer *et al*. 2011). Poor stand establishment caused by poor germination, delayed emergence and/or slow growth is a key bottleneck for expansion of sorghum in the northern hemisphere, despite its relatively hardy response to several other abiotic stresses (Knoll *et al*. 2008). There is growing evidence for discrete genetic control of germination, emergence and seedling development of sorghum under chilling conditions (e.g. Balota *et al*. 2010). We found an interdependency of several biomass and emergence traits, whereas germination was found to cluster in another group together with biomass traits under optimum temperature conditions. Similar contradictory results were reported by Pedersen & Toy (2001) and Balota *et al*. (2010) in tests of germination and emergence in sorghum and maize. We found stronger correlations between different locations and sowing times than between germination and field emergence.

Broad variability for germination and emergence under low temperature conditions is a great advantage in breeding sorghum for adaptation to temperate regions with potentially low early sowing temperatures (Knoll *et al*. 2008; Burow *et al*. 2011). Differences in germination rates between the two seed lots we used can be attributed to environmental conditions during the seed production (Mohamed, Clark & Ong 1985; Harris, Hamdi & Oda 1987; Blönder *et al*. 2007; Elwell *et al*. 2011). Seed weight was found not to influence cold emergence; however, reflecting previous results in sorghum and maize (Maranville & Clegg 1977; Hund *et al*. 2004). On the other hand, the number of days to onset of cold germination was found to significantly influence emergence in soil.

Retarded seedling establishment and death under chilling is associated with root growth suppression

In thermophilic plants like sorghum low temperature causes water deficit by reducing water uptake without the necessary reduction of leaf transpiration rate (Aroca *et al*. 2001; Aroca, Porcel & Ruiz-Lozano 2012). In temperate climates cold-sensitive crops are sown during spring time, and it is not uncommon for seeds to encounter suboptimal soil temperatures that can potentially influence the developing root system and consequent water and mineral acquisition (Bloom *et al*. 2004; Huang *et al*. 2005). Time lapse studies of root elongation rate in sorghum seedlings demonstrated that the rate of root elongation is determined primarily by temperature rather than by the diurnal cycle (Iijima *et al*. 1998).

Chilling-stressed plants typically show wilting phenotypes similar to drought symptoms; hence it was unsurprising that we found many relationships of chilling survival traits to root traits. There are several reports emphasizing the role of root development on chilling survival. Balota *et al*. (2010) demonstrated that root biomass is stable in chilling stress-tolerant genotypes, while Menkir & Larter (1987) showed that seedling vigour under field conditions was correlated to root

biomass in controlled conditions. In maize and rice roots, a chilling-induced reduction in conductance and a corresponding decrease in leaf water potential were found to cause a lethal decline in leaf relative water content (Melkonian, Yu & Setter 2004; Matsumoto *et al*. 2009). Besides the hydraulic conductance, the root architecture is also highly relevant for temperature responses. In maize, for example, Hund *et al*. (2008) found that the primary lateral root length was related to photosynthetic activity, chlorophyll content values and seedling DW under chilling conditions, while the axial root length was more important under optimum temperatures. Knipfer & Fricke (2011) demonstrated that much of the water uptake by 13–17-day-old barley seedlings is achieved by seminal and lateral roots rather than axial adventitious roots. Hence, the increased lateral root development of M71 could contribute to its greater tolerance to extended physiological maintenance under chilling stress. Detailed analysis of root hydraulic conductance and transpiration rate differences should give novel insight into the root contribution to chilling tolerance and the genetic mechanisms underlying QTL regions implicated in cold survival, chlorophyll content, root developmental rate and root biomass.

There is growing interest from both crop breeders and plant physiologists in the development and architecture of sorghum roots, because of the vast variability present in sorghum and its substantial root depth and simpler root architecture compared with other cereals (Singh *et al*. 2011). Root architecture traits such as nodal root angle were found by Mace *et al*. (2012) to underlie an array of complex traits related to the stay-green phenotype and to seed yield in sorghum. The stay-green trait gives sorghum plants the ability to withstand post-flowering drought stress by keeping their flag-leaves green for up to 60 days after flowering, consequently increasing grain yield under stress conditions (Borrell, Hammer & Van Oosterom 2001; Harris *et al*. 2007). Although the physiological basis remains to be understood, similar stay-green loci were reported to be important during early-stage drought (Kassahun *et al*. 2010). Interestingly, some of the QTL for root traits and chlorophyll content that we identified in the present study correspond to positions of stay-green loci identified in earlier studies (Harris *et al*. 2007; Mace *et al*. 2012), for example Stg2 on Sb03. A combination of these loci with QTL conferring other chilling-tolerance mechanisms may facilitate the development of sorghum varieties with generally improved root architecture and a broad abiotic stress tolerance.

Survival is an effective indicator of early-stage chilling tolerance

Emergence in cold soil is not useful if the seedling cannot subsequently survive the chilling conditions. Thermophilic C4 grasses like maize and sorghum are especially sensitive at the early stage of the autotrophic phase (Bhosale *et al*. 2007). Unlike short-term chilling stress, growth under sustained chilling can cause irreversible damage to the photosynthetic apparatus (Nie & Baker 1991; Leipner, Fracheboud & Stamp 1997), leading to interrupted carbon assimilation and death. When not coupled with appropriate stomatal closure, reduced water and nutrient uptake as a result of prolonged root chilling can lead to wilting and death (Aroca *et al*. 2001, Huang *et al*. 2005). We found transgressive segregation and high heritability for survival under prolonged chilling stress, indicating a complex genetic determination. Cold treatment at different time points after emergence revealed highest sensitivity to chilling at 4–5 days after emergence (Fig. 3). Chilling stress beginning at sowing reorganizes the metabolic processes of the seedling more than stress after emergence, where the plant is suddenly subjected to suboptimal temperatures without hardening and can suffer membrane damage in the roots and other organs (Ahamed *et al*. 2012; Aroca *et al*. 2012). Maize plants were found to be capable of acclimatization to cold stress down to 4 °C after exposure to cool temperatures of 14 °C (Prasad *et al*. 1994). This kind of chilling acclimation in maize reduces photooxidation of leaves in subsequent phases of chilling stress (Leipner, Fracheboud & Stamp 1997). Similar acclimation of rice seedlings via *DREB*-induced acclimatization was recently reported by Mao & Chen (2012).

Because of its correlation to both emergence and biomass under chilling stress, cold survival is suggested as the best surrogate trait in order to capture and combine variability for both early-stage emergence and seedling development under chilling conditions. Chilling survival has been used in soybean, maize and rice chilling stress studies (Prasad, Anderson & Stewart 1994; Yun *et al*. 2009). In temperate climates, it is possible to a certain extent to ensure that sowing occurs under favourable soil temperature conditions; however, the unpredictability of post-emergence temperatures (e.g. Fig. 3) makes chilling tolerance at 4–5 days after emergence a vital trait for breeding. We identified cold survival as a useful indicator trait, which encompasses the genetic potential of a genotype for the genetically and developmentally different determinants of germination, emergence and seedling vigour under chilling stress in field conditions. Dissection of the main QTL involved in the expression of these different factors will help identify the underlying genes and improve our understanding of the multifaceted physiology of the stress response in the most chilling-tolerant breeding lines.

QTL co-localization hubs may control chilling responses on multiple regulatory levels

The significant QTL co-localization hotspots we identified represent putative regulatory regions involved in maintenance of general seedling vigour under both optimum and stress conditions. The hotspot of main-effect and epistatic interaction QTL on chromosome Sb06 represents a new source of chilling tolerance for sorghum; previous cold tolerance studies in Chinese Kaoliang germplasm and other African materials identified important QTL only on chromosomes Sb01, Sb02, Sb03, Sb04 and Sb10 (Knoll *et al*. 2008; Burow *et al*. 2011). On the other hand, this region of Sb06 has also been implicated in ergot resistance, drought tolerance and sugar metabolism (Prah *et al*. 2008; Mace & Jordan 2011; Shiringani & Friedt 2011). A cluster of agronomic

trait QTL under contrasting photoperiod conditions were recently mapped on Sb06 using a sequencing-based mapping approach (Zou *et al*. 2012). These results may indicate a general role of this region in sorghum growth and development under various environmental conditions, making it an interesting focus for studies of selection and fitness during sorghum evolution.

Diverse and potentially complementary regulatory systems were identified among potential candidate genes within the major chilling stress QTL hotspots. The QTL hot spot support interval harbours interesting hormone catabolism and hormone responsive candidate genes, including a cytokinin oxidase/dehydrogenase (Sb06g022930), a cytokinin-induced response regulator gene (Sb06g022960), auxin response protein or auxin response factor-like genes (Sb06g022720 and Sb06g022810) and ABA responsive-like genes (Sb06g023170 and Sb06g023180). This suggests that the observed QTL co-localization for root development, emergence and survival phenotypes might be the result of hormone-driven regulation of different developmental processes and abiotic stress responses. Further gene expression and metabolite profiling studies on seedlings grown on exogenous hormone-supplemented media, under chilling and optimum conditions, could provide additional complementary data to elucidate the molecular basis of seedling development under chilling stress conditions in sorghum.

Besides hormone-responsive elements, with roles, for example in root architecture, we also found genes related to membrane transport, chloroplast stability, chitin catabolism and cold adaptation, all of which have potential roles in different chilling responses. The transgressive segregation for chilling survival and seedling development in the RIL population presumably arises from a combination of complementary seedling development and chilling tolerance regulating genes from the two parental lines. This demonstrates the high genetic complexity as well as the potential to improve early stage seedling development under temperate conditions.

CONCLUSIONS

Despite the highly complex nature of early stage chilling stress responses, the presence of major QTL co-localization hot spots and highly heritable surrogate traits can potentially enhance breeding for early stage chilling tolerance in sorghum, via marker-assisted selection in combination with precision phenotyping. Association genetics approaches using higher marker densities in genetically diverse populations will help to fine-map major QTL hubs and identify further useful variation for abiotic stress tolerance. Detailed physiological studies to dissect the effects of chilling stress on root conductance, shoot growth and seedling development rates can contribute significantly to understanding the exceptional abiotic stress tolerance of sorghum. Systems biological approaches combining genetic, metabolic and molecular physiological studies could provide important new insight for breeding of resistant crops to meet growing world demand for food and fuel in the face of climatic unpredictability and increasing abiotic stress constraints.

ACKNOWLEDGMENTS

The authors thank Arndt Zaccharias (KWS Saat AG, Einbeck, Germany) for seed production in Italy and Chile. Excellent technical assistance was provided by Swetlana Renner, Nelly Weiss, Mario Tolksdorf, Malte Luh, Markus Kolmer, David Wiese and Helge Flüss. This research was funded by the BMBF *Bioenergy 2021* grant 0315421B.

REFERENCES

Ahamed A., Murai-Hatano M., Ishikawa-Sakurai Hayashi H., Kawamura Y. & Uemura M. (2012) Cold stress-induced acclimation in rice is mediated by root-specific aquaporins. *Plant & Cell Physiology* **53,** 1445–1456.

Andaya V.C. & Mackill D.J. (2003) Mapping of QTLs associated with cold tolerance during the vegetative stage in rice. *Journal of Experimental Botany* **54,** 2579–2589.

Aroca R., Tognoni F., Irigoyen J.J., Sánchez-Diaz M. & Pardossi A. (2001) Different root low temperature response of two maize genotypes differing in chilling sensitivity. *Plant Physiology & Biochemistry* **39,** 1067–1073.

Aroca R., Porcel R. & Ruiz-Lozano J.M. (2012) Regulation of root water uptake under abiotic stress conditions. *Journal of Experimental Botany* **63,** 43–57.

Balota M., Payne W.A., Veeragoni S.K., Stewart B.A. & Rosenow D.T. (2010) Respiration and its relationship to germination, emergence, and early growth under cool temperatures in sorghum. *Crop Science* **50,** 1414–1422.

Bhosale S.U., Rymen B., Beemster T.S., Melchinger A.E. & Reif J.C. (2007) Chilling tolerance of central European maize lines and their factorial crosses. *Annals of Botany* **100,** 1315–1321.

Blönder C., Goebel C., Feussner I., Gatz C. & Polle A. (2007) Warm and cold parental reproductive environments affect seed properties, fitness, and cold responsiveness in *Arabidopsis thaliana* progenies. *Plant, Cell & Environment* **30,** 165–175.

Bloom A.J., Zweienicki M.A., Passioura J.B., Randall L.B., Holbrook N.M. & Clair A.S. (2004) Water relations under root chilling in a sensitive and tolerant tomato species. *Plant, Cell & Environment* **27,** 971–979.

Borrell A., Hammer G. & Van Oosterom E. (2001) Stay-green: a consequence of the balance between supply and demand for nitrogen during grain filling? *Annals of Applied Biology* **138,** 91–95.

Brar G.S., Steiner J.L., Unger P.W. & Prihar S.S. (1992) Modelling sorghum seedling establishment from soil wetness and temperature of drying seed zones. *Agronomy Journal* **84,** 905–910.

Burow G., Burke J.J., Xin Z. & Franks C.D. (2011) Genetic dissection of early-season cold tolerance in sorghum (*Sorghum bicolor* (L.) Moench. *Molecular Breeding* **28,** 391–402.

Chinnusamy V., Zhu J. & Zhu J.K. (2007) Cold stress regulation of gene expression in plants. *Trends in Plant Science* **12,** 444–451.

Du Z., Ling Y., Zhang Z. & Su Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Research* **38,** 1–7.

Edwards E.J. & Smith S.A. (2010) Phylogenetic analyses reveal the shady history of C4 grasses. *Proceedings of the National Academy of Sciences of the United States of America* **107,** 2532–2537.

Elwell A.L., Gronwall D.S., Miller N.D., Spalding E.P. & Brooks T.L. (2011) Separating parental environment from seed size effects on next generation growth and development in Arabidopsis. *Plant, Cell & Environment* **34,** 291–301.

Fiedler K., Bekele W.A., Friedt W., Snowdon R., Stützel H., Zacharias A. & Uptmoor R. (2012) Genetic dissection of the temperature dependent emergence processes in sorghum using a cumulative emergence model and stability parameters. *Theoretical and Applied Genetics* **125,** 1647–1661.

Harris D., Hamdi Q.A. & Oda A.C.T. (1987) Germination and emergence of *Sorghum bicolor* genotypic and environmentally induced variation in the response to temperature and depth sowing. *Plant, Cell and Environment* **10,** 501–508.

Harris K., Subudhi P.K., Borrell A., Jordan D., Rosenow D., Nguyen H., Klein R. & Mullet J. (2007) Sorghum stay-green QTL individually reduce post-flowering drought induced senescence. *Journal of Experimental Botany* **58,** 327–338.

Huang L., Ye Z., Bell R.W. & Dell B. (2005) Boron nutrition and chilling tolerance of warm climate crop species. *Annals of Botany* **96,** 755–767.

Hund A., Frachebound Y., Soldati A. & Frascaroli E. (2004) QTL controlling root and shoot traits of maize seedlings under cold stress *Theoretical and Applied*. *Genetics* **109,** 618–629.
Hund A., Fracheboud Y., Soldati A. & Stamp P. (2008) Cold tolerance of maize seedlings as determined by root morphology and photosynthetic traits. *European Journal of Agronomy* **28,** 178–185.
Iijima M., Oribe Y., Horibe Y. & Kono Y. (1998) Time lapse analysis of root elongation rates of rice and sorghum during the day and night. *Annals of Botany* **81,** 603–607.
Kassahun B., Bidinger F.R., Hash C.T. & Kuruvinashetti M.S. (2010) Stay-green expression in early generation sorghum (*Sorghum bicolor* (L.) Moench) QTL introgression lines. *Euphytica* **172,** 351–362.
Knipfer T. & Fricke W. (2011) Water uptake by seminal and adventitious roots in relation to whole-plant water flow in barley (*Hordeum vulgare* L.). *Journal of Experimental Botany* **62,** 717–733.
Knoll J., Gunaratna N. & Ejeta G. (2008) QTL analysis of early-season cold tolerance in sorghum. *Theoretical and Appied Genetics* **116,** 577–587.
Krzywinski M., Schein J.E., Birol I., Connors J., Gascoyne R., Horsman D., Jones S.J. & Marra M.A. (2009) CIRCOS: an information aesthetic for comparative genomics. *Genome Research* **19,** 1639–1645.
Lee S.H., Singh A.P., Chung G.C., Kim Y.S. & Kong I.B. (2002) Chilling root temperature causes rapid ultrastructural changes in cortical cells of cucumber (*Cucumis sativus* L.) root tips. *Journal of Experimental Botany* **53,** 2225–2237.
Leipner J., Fracheboud Y. & Stamp P. (1997) Acclimation by suboptimal growth temperature diminishes photooxidative damage in maize leaves. *Plant, Cell & Environment* **20,** 366–372.
Mace E.S. & Jordan D.R. (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of gene-rich regions with significant implications for crop improvement. *Theoretical and Applied Genetics* **123,** 169–191.
Mace E.S., Singh V., Van Oosterom E.J., Hammer G.L., Hunt C.H. & Jordan D.R. (2012) QTL for nodal root angle in sorghum (*Sorghum bicolor* L.Moench) co-locate with QTL for traits associated with drought adaptation. *Theoretical and Applied Genetics* **124,** 97–109.
Mao D. & Chen C. (2012) Colinearity and similar expression pattern of rice *DREB1s* reveal their functional conservation in the cold-responsive pathway. *PLoS ONE* **7,** e47275.
Maranville J.W. & Clegg M.D. (1977) Influence of seed size and density on germination, seedling emergence and grain yield in grain sorghum. *Agronomy Journal* **69,** 329–330.
Matsumoto T., Lian H.L., Su W.A., Tanaka D., Liu C., Iwasaka I. & Kitagwa Y. (2009) Role of the Aquaporin in PIP1 subfamily in the chilling tolerance of rice. *Plant & Cell Physiology* **50,** 216–229.
Melkonian J., Yu L.X. & Setter T.L. (2004) Chilling responses of maize (*Zea mays* L.) seedlings: root hydraulic conductance, absicisic acid and stomatal conductance. *Journal of Experimental Botany* **55,** 1751–1760.
Menkir A. & Larter E.N. (1987) Emergence and seedling growth of inbred lines of corn at suboptimal root-zone temperatures. *Canadian Journal of Plant Sciences* **67,** 409–415.
Mercer K., Alexander H.M. & Snow A.A. (2011) Selection on seedling emergence timing and size in an annual plant, *Helianthus annuus* (common sunflower, Asteraceae). *American Journal of Botany* **98,** 975–985.
Mohamed H.A., Clark J.A. & Ong C.K. (1985) The influence of temperature during seed development on the germination characteristics of millet seeds. *Plant, Cell & Environment* **8,** 361–362.
Nie G.Y. & Baker N.R. (1991) Modification of thylakoid composition during development of maize leaves at low growth temperatures. *Plant Physiology* **95,** 184–191.
Oyediran I.O., Hibbard B.E. & Clark T.L. (2004) Prairie grasses as hosts of the western corn rootworm (Coleoptera: Chrysomelidae). *Environmental Entomology* **33,** 740–747.
Patanè C., Cavallaro V., Avola G. & D'Agosta G. (2006) Seed respiration of sorghum (*Sorghum bicolor* (L.) Moench). *Seed Scicence Research* **16,** 251–260.
Paterson A., Bowers J.E., Bruggmann R., *et al.* (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457,** 551–556.
Payne W.A., Batata M. & Rosenow D.T. (2003) Sorghum diversity for germination and coleoptile elongation under cool conditions. *International Sorghum and Millets Newsletter* **44,** 76–79.
Peacock J.M. (1982) Response and tolerance of sorghum to temperature stress. Sorghum in the eighties: *Proceedings of the International Symposium on Sorghum*, 2–7 November 1981, Patancheru. A.P. India.
Pedersen J.F. & Toy J.J. (2001) Germination, emergence, and yield of 20 plant-color, seed-color near-isogenic lines of grain sorghum. *Crop Science* **41,** 107–110.
Prah D.K., Jorda D.R., Aitken A.B., Mace E.S., Jun-ai P., McIntyee C.L. & Godwin I.D. (2008) QTL analysis of ergot resistance in sorghum. *Theoretical and Applied Genetics* **117,** 369–382.
Prasad T.K., Anderson M.D., Martin B.A. & Stewart C.R. (1994) Evidence for chilling-induced oxidative stress in maize nd a regulatory role for hydrogen peroxide. *The Plant Cell* **6,** 65–74.
Prasad T.K., Anderson M.D. & Stewart C.R. (1994) Acclimation, hydrogen peroxide and abscic acid protect mitochondria against irreversible chilling injury in maize seedlings. *Plant Physiology* **105,** 619–627.
Reddy B.V.S., Ramesh S., Kumar A.A., Wani S.P., Ortiz R., Ceballos H. & Sreedevi T.K. (2008) Bio-fuel crops research for energy security and rural development in developing countries. *Bioenergy Research* **1,** 248–258.
Seo J. & Shneiderman B. (2002) Interactively exploring hierarchical clustering results. *Computer* **35,** 80–86.
Shiringani A.L. & Friedt W. (2011) QTL for fibre-related traits in grain × sweet sorghum as a tool for the enhancement of sorghum as a biomass crop. *Theoretical and Applied Genetics* **123,** 999–1011.
Shiringani A.L., Frisch M. & Friedt W. (2010) Genetic mapping of QTLs for sugar-related traits in a RIL populaiton of *Sorghum bicolor* L. Moench. *Theoretical and Applied Genetics* **121,** 323–336.
Singh V., Van Oosterom E.J., Jordan D.R., Hunt C.H. & Hammer G.L. (2011) Genetic variability and control of nodal root angle in sorghum. *Crop Science* **51,** 2011–2020.
Subbarao G.V., Rondon M., Ito O., Ishikawa T., Rao I.M., Nakahara K., Lascano C. & Berry W.L. (2006) Biological nitrification inhibition (BNI) – is it a widespread phenomenon? *Plant and Soil* **294,** 5–18.
Subbarao G.V., Zakir H.A.K.M., Nakahara K., *et al.* (2009) Biological nitrification inhibition (BNI) potential in sorghum. *Proceedings of the International Plant Nutrition Colloquium XVI*, Department of Plant Sciences, UC Davis, USA.
Utz H.F. & Melchinger A.E. (1996) PLABQTL: a program for composite interval mapping of QTL. *Journal of Agricultural Genomics* **2,** 1–5.
Yu J., Tuinistra M.R., Classsen M.M., Gordon W.B. & Witt M.D. (2003) Analysis of cold tolerance in sorghum under controlled environment conditions. *Field Crops Research* **86,** 21–30.
Yun K.Y., Park M.R., Mohanty B., Herath V., Xu F., Mauleon R., Wijaya E., Bajic V.B., Bruskiewich R. & Reyes B.G. (2009) Transcriptional regulatory network triggered by oxidative signals configures the early response mechanisms of japonica rice to chilling stress. *BMC Plant Biology* **10,** 16.
Zhi-Hong Z., Li S., Wei L., Wei C. & Ying-Guo Z. (2005) A major QTL conferring cold tolerance at the early seedling stage using recombinant inbred lines of rice (*Oryza sativa* L.). *Plant Science* **168,** 527–534.
Zou G., Zhai G., Feng Q., *et al.* (2012) Identification of QTLs for eight agronomically important traits using an ultra-throughput sequencing in sorghum under contrasting photoperiods. *Journal of Experimental Botany* **63,** 5451–5462.

Received 7 September 2012; received in revised form 15 August 2013; accepted for publication 17 August 2013

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Supplementary Methods

Figure S1. Soil temperature at 5 cm below soil surface during sowing and emergence of the 2009 cropping season experiments at three locations Gross-Gerau, Giessn and Hannover and two sowing times.

Figure S2. Root architecture of SS79 and M71 plants grown under chilling conditions (13 °C) for 2 weeks followed by 1 week recovery at 25 °C/20 °C day/night.

Figure S3. Circular genetic map from the *Sorghum bicolor* cross SS79 × M71 showing positions of main-effect QTL and two-locus epistatic QTL (internal connecting lines) for all

germination, greenhouse and field experiments conducted with the 2008 seed lot from Gross Gerau. The light grey tracks represent traits measured after chilling at sowing (CAS), blue bars and connecting lines represent QTL and epistatic interactions for germination and emergence related traits; orange: field emergence; black: root traits; red: survival; yellow: shoot traits; green: leaf chlorophyll traits. A full list of the specific traits numbered from 1 (innermost track) to 63 (outermost track) is included. The blue and green highlights show regions with at least two years overlapping QTL. The green colour shows QTL colocalizations including main effect QTL during the validation experiment.

Table S1. An excel file with three different sheets showing the traits Pearson's correlations for the experiments conducted from 2009 to 2011.

Tables S2–S4. Detailed main effect and epistatic QTL detected. Tables 2a, 3a and 4a show the main effect QTL for year 2009, 2010 and 2011 and 2b, 3b and 4b show the epistatic QTL.

Table S5. List of positional and functional candidate genes identified from QTL co-localization hot spots.

3 High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array

Bekele WA, Wieckhorst S, Friedt W, Snowdon RJ

Plant Biotechnology Journal 2013

Vol 11, pp. 1112-1125

Plant Biotechnology Journal

Plant Biotechnology Journal (2013) **11**, pp. 1112–1125 doi: 10.1111/pbi.12106

High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array

Wubishet A. Bekele[1], Silke Wieckhorst[2], Wolfgang Friedt[1] and Rod J. Snowdon[1,*]

[1]*Department of Plant Breeding, Justus Liebig University, Giessen, Germany*
[2]*KWS Saat AG, Einbeck, Germany*

Received 30 April 2013;
revised 28 June 2013;
accepted 9 July 2013.
**Correspondence* (Tel +49 641 9937423;
fax +49 641 9937429;
email rod.snowdon@agrar.uni-giessen.de)

Keywords: Illumina, Infinium, single-nucleotide polymorphism, SNP chip.

Summary

With its small, diploid and completely sequenced genome, sorghum (*Sorghum bicolor* L. Moench) is highly amenable to genomics-based breeding approaches. Here, we describe the development and testing of a robust single-nucleotide polymorphism (SNP) array platform that enables polymorphism screening for genome-wide and trait-linked polymorphisms in genetically diverse *S. bicolor* populations. Whole-genome sequences with 6× to 12× coverage from five genetically diverse *S. bicolor* genotypes, including three sweet sorghums and two grain sorghums, were aligned to the sorghum reference genome. From over 1 million high-quality SNPs, we selected 2124 Infinium Type II SNPs that were informative in all six source genomes, gave an optimal Assay Design Tool (ADT) score, had allele frequencies of 50% in the six genotypes and were evenly spaced throughout the *S. bicolor* genome. Furthermore, by phenotype-based pool sequencing, we selected an additional 876 SNPs with a phenotypic association to early-stage chilling tolerance, a key trait for European sorghum breeding. The 3000 attempted bead types were used to populate half of a dual-species Illumina iSelect SNP array. The array was tested using 564 *Sorghum* spp. genotypes, including offspring from four unrelated recombinant inbred line (RIL) and F_2 populations and a genetic diversity collection. A high call rate of over 80% enabled validation of 2620 robust and polymorphic sorghum SNPs, underlining the efficiency of the array development scheme for whole-genome SNP selection and screening, with diverse applications including genetic mapping, genome-wide association studies and genomic selection.

Introduction

Sorghum is widely grown as a staple cereal crop, particularly in Africa and parts of Asia but also for various uses in other parts of the world. In Australia and the US Southern Plains, grain sorghum represents a drought-tolerant alternative to maize production for livestock feeding, while in Europe, China and North America, interest is also growing rapidly in the use of sweet and/or grain sorghum forms as a potentially drought-tolerant and nutrient-efficient alternative to maize for bio-energy production (Rooney *et al.*, 2007). Breeding sorghum for temperate regions necessitates the combination of chilling tolerance from appropriate germplasm resources with photoperiod adaptation and other appropriate agronomic characters (e.g. high dry-matter biomass or grain yields), whereas in subsistent arid and semi-arid farming systems, grain yield and seed quality are the most vital traits. In comparison with more established crops like maize, wheat and barley, breeding of sorghum for bio-energy and livestock feeding is a relatively young enterprise, and an enormous genetic potential for improvement in the crop has yet to be tapped by breeders. The ability to intercross cultivated sorghum races with related subspecies for expansion of genetic diversity and improvement in key traits is a unique aspect of this crop (Washburn *et al.*, 2013).

The small diploid genome of sorghum, the availability of a completed reference genome sequence (Paterson *et al.*, 2009) and the consequent ability to develop cost-effective, high-throughput tools for whole-genome screening make sorghum a strikingly amenable crop for the application of genomics-based breeding methods. In particular, the dramatically falling costs of genome-wide screening for single-nucleotide polymorphisms (SNPs), using high-density SNP array technologies (see Batley and Edwards, 2007) or genotyping-by-sequencing (Chia and Ware, 2011; Davey *et al.*, 2011; Elshire *et al.*, 2011; Morris *et al.*, 2013) on next-generation sequencing platforms (Metzker, 2010), has opened the way for genomic selection (Jannink *et al.*, 2010) or predictive breeding strategies (Riedelsheimer *et al.*, 2012). Such techniques have the potential to considerably accelerate selection gain and improve the effectiveness of breeding.

Nelson *et al.* (2011) used reduced-representation sequencing of restriction-site-associated DNA (RAD; see Baird *et al.*, 2008) to discover genome-wide SNPs in a panel of 8 genetically diverse grain sorghum genotypes. By whole-genome resequencing of two sweet sorghum accessions and one grain sorghum, Zheng *et al.* (2011) not only detected over 1 million genomic SNPs, but also demonstrated that over 1500 genes differentiate between sweet and grain sorghum. This gene diversity spans important processes of high relevance for breeding, for example sugar and starch metabolism, lignin biosynthesis and stress responses. This demonstrates that a broad genetic basis of genome-wide sequence variation is necessary to capture SNP variation of general relevance for genetic diversity and breeding towards different end-use scenarios. In the present study, we supplemented the sequence data from Zheng *et al.* (2011) by resequencing one further sweet sorghum and one further grain sorghum genotype, both of geographically and genetically

divergent origin. Furthermore, we generated two phenotypic pools of 30 recombinant inbred lines (RILs) each, segregating strongly for early-stage chilling tolerance, from a cross between the latter two genotypes. These 60 RILs were skim-sequenced with a moderate genome coverage to identify trait-linked SNPs of relevance to European energy sorghum breeding. Using six whole-genome sequences (including the *S. bicolor* reference sequence from the grain sorghum BTx623), we were able to detect an extremely high number of high-quality genome-wide SNPs with high allele frequencies in genetically diverse *S. bicolor* germplasm, encompassing both grain and sweet sorghum types. The SNP array we developed using this data (Figure 1) was used to validate 2620 robust and polymorphic SNPs in a panel of 564 genetically diverse sorghum accessions, including four unrelated RIL and $F_{2\text{-}3}$ mapping populations and a *Sorghum* spp. genetic diversity collection.

Results and discussion

Paired-end short sequence mapping

The technological advancement of paired-end sequencing makes mapping of short DNA sequence reads onto the sorghum genome extremely precise and efficient despite the large proportion of repeat units (Paterson *et al.*, 2009; Zheng *et al.*, 2011). In contrast, Nelson *et al.* (2011) applied single-end Illumina sequencing for SNP detection sequencing in eight diverse sorghum lines. Our resequencing strategy mapped more than 90% (Table 1) of the total reads onto the reference genome, comparable to the frequency of successfully mapped reads reported by Zheng *et al.* (2011); with single-end reads from restriction-site-associated DNA (RAD) sequences only 30% of the reads could be correctly mapped by Nelson *et al.* (2011).

Resequencing in sweet and grain sorghums

The huge quantity of SNPs we detected in SS79, M71, Keller, E-tian and Ji-2731 underlines the high level of polymorphism present in sorghum. In the study of Zheng *et al.* (2011), Ji-2731 was reported to show the largest SNP diversity compared with BTx623; however, we revealed that many of the identified polymorphisms were heterozygote state SNPs. The patterns of SNP distribution we observed across the genome reflected other reports on sorghum and other plants (Morris *et al.*, 2013; Paterson *et al.*, 2009; Zheng *et al.*, 2011). Low SNP densities were seen in centromeric regions, with high numbers of SNPs towards the chromosome ends (Figure 2). As expected, the sweet sorghum genotypes SS79 and Keller exhibited greater SNP diversity compared with the grain sorghum reference sequence than was observed in the other grain sorghum genotypes (Table 2). Interestingly, significant genome-scale difference between the three sweet sorghums and the three grain sorghum types (including Btx623) was observed in a large chromosome block on chromosome Sb10 (25-45 Mb) (Figure 2). On the other hand, we observed a region of approximately 25 Mbp on chromosome Sb02 with low polymorphism in all five accessions compared with Btx623, while a 35 Mbp region on chromosome Sb04 was strongly divergent in the two grain sorghum genotypes we analysed compared with the three sweet sorghums and the Btx623 grain sorghum reference. More detailed study of such regions can potentially give interesting insights into sorghum domestication and adaptation to diverse climatic and agricultural systems. For example, quantitative trait loci (QTL) influencing sorghum maturity or photoperiod sensitivity, and with a pleiotropic effect on sugar, biomass and grain yield, have been identified in biparental sweet x grain sorghum populations within the corresponding region of chromosome Sb04 (Murray *et al.*, 2008; Shiringani *et al.*, 2010). In a nested association mapping study using more than 1000 individuals from 24 families, Mace and Jordan (2013) identified 40 small-effect QTL showing synteny to flowering-time QTL in maize. Their results revealed hotspots for flowering-time QTL on chromosomes Sb03 and Sb04. Some of the candidate genes they reported lie within the interesting region of differentiation that we identified. The most prominent of these is Sb04g008320 (*SbFT*; 9.47Mbp), a

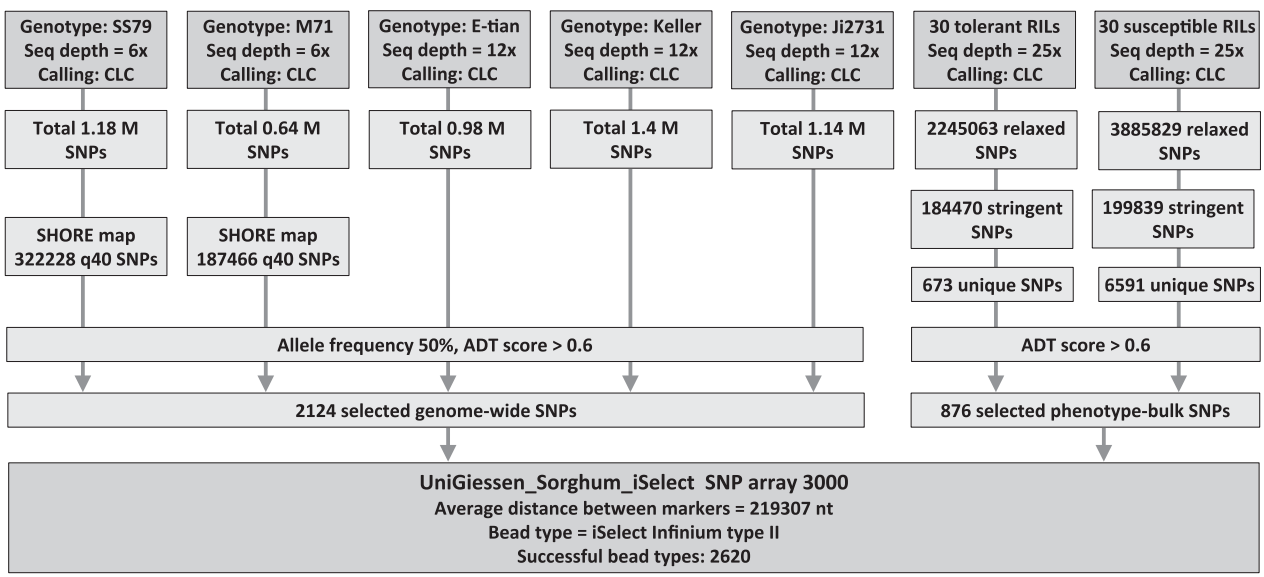


Figure 1 Outline of the single-nucleotide polymorphism (SNP) detection, filtering and array development scheme. Single-nucleotide polymorphism calling was performed with CLC Genomics Workbench and SHOREmap, respectively. Details of the phenotypic bulk SNP identification are described in Experimental procedures.

Table 1 Summary of SNPs and other variations detected using CLC genomics between the parental lines SS79 and M71, and between two phenotypic pools of 30 SS79 × M71 RILs each, with high tolerance and high sensitivity to early-stage chilling stress, respectively

Parental lines in comparison with BTx623	SS79	M71
Number of reads after trim	48 712 177	44 167 614
Total mapped reads	48 512 190	40 586 315
Fraction of reference covered	0.89	0.88
Average coverage excluding zero coverage regions	6.51	5.78
SNPs in genic regions	133 094	112 185
SNPs in noncoding regions	1 603 454	778 829
Total bi-allelic SNPs	1 184 364	636 307
Bi-allelic SNPs/kb	1.6	0.86
Deletions	195 165	220 776
Amplifications	114 772	104 363
Phenotype pools in comparison with BTx623	**Chilling-tolerant pool**	**Chilling-sensitive pool**
Total mapped reads	170 983 983	189 267 873
Fraction of reference covered	0.89	0.90
Average coverage excluding zero coverage regions	23.16	25.62
SNPs identified with relaxed settings	2 245 063	3 885 829
SNPs identified with stringent settings	184 470	1 999 839
SNPs unique to the phenotypic pool	673	7264
Unique SNPs in genic regions	84	831

homologue of Arabidopsis *Flowering-time Locus T* (AT1G65480; *FLT*).

Chromosomes Sb04, Sb06 and Sb09 are known hotspots of differentiation between sweet and grain sorghum (Zheng *et al.*, 2011). The differentiation region in this study includes large deletions and copy-number variations in vital genes like the RNA polymerase subunit genes Sb04g009441 and Sb04g009 491. Another interesting example is the gene Sb04g035450, which was absent in the two sweet sorghum genotypes Keller and Etian (Zheng *et al.*, 2011). Sb04g035450 is annotated as an ionotropic glutamate receptor (GLR) related gene and has homology to Arabidopsis *GLR* genes involved in calcium ion transport, light stimulus response and regulation of C:N ratio and metabolism (Kang and Turano, 2003). The presence of significant and widespread structural variation in sweet sorghum genomes compared with the grain sorghum reference sequence underlines the need for a dedicated, *de novo* sweet sorghum reference genome for future genome research. This will help to better understand the genetic basis of sweet and grain sorghum differentiation. It will also provide direct access to gene sequences and regulatory regions that are specific to energy-related traits in sweet sorghum, but not necessarily assembled or annotated in the grain sorghum reference genome.

Comparison of SNPs detected by CLC Genomics Workbench and SHOREmap in the two parental genotypes SS79 and M71 showed that more than 90% of all SNPs called by SHOREmap were also detected by CLC, whereby only 70% of SHOREmap Q25 SNPs were also detected by CLC Genomics Workbench. For downstream selection and array development, we only used SNPs that were called by both programmes with quality scores of at least Q40. A total of 163,027 SNPs were selected that had quality scores ≥ Q40 in SS79 and M71 and were also called in all six genotypes with an allele frequency of 50% (Figure 1).

QTL delineation with phenotype-bulk sequence analysis

Sequence mapping in the two phenotype-based pools achieved 24-fold (24×) average coverage in each group (Table 1). Single-nucleotide polymorphism detection with both stringent and more relaxed parameters showed that the early-stage chilling-susceptible pool has more intrapool polymorphism than the chilling-resistant pool. Under relaxed conditions, a total of 3 885 829 SNPs were detected within the sensitive pool, while under stringent settings this number was reduced to 1 999 839 SNPs. In contrast, under relaxed settings, 2 245 063 SNPs were detected in the susceptible pool, and only 184 470 were detected under stringent calling settings in the chilling-resistant pool. However, despite the extremely large number of SNPs detected in each phenotype pool, only 7264 were found to be unique to one or other of the pools. Furthermore, the chilling-resistant pool contained only 673 unique SNPs, representing only 0.36% of the total number of SNPs (Figure 1). Somewhat more SNPs (6591 or 3.2%) were detected that were unique within the susceptible pool. The extremely low number of unique SNPs associated with the phenotypic pool representing RILs with high chilling stress resistance demonstrates the enormous power of bulked-segregant genome-wide sequencing for detection of markers tightly linked to QTL for agronomic traits in crops. A similar bulked segregant analysis was used by Hu *et al.* (2012) to identify a gene for pod shattering in rapeseed, for example.

The 673 unique SNPs associated with chilling resistance were found to be dispersed across the whole-genome; however, specific regions of chromosomes 2, 4, 5 and 8 contained more than 25 SNPs/10 Mbp that were unique to the chilling stress phenotype. Interestingly, 12.9% of the SNPs unique to the two phenotypic pools were found to be genic markers, although the SNP selection was not biased towards coding sequences. A total of 831 genic SNPs were unique to the chilling-susceptible pool and 84 to the chilling-resistant pool. A GO-term annotation enrichment test revealed enrichment of genes involved in death, growth response to stimuli and antioxidant activity.

Selection of markers for array development

Flanking 60 nt sequences for 76 574 SNPs were uploaded into the Illumina Assay Design Tool for final selection of SNPs to be included on the iSelect BeadChip. From 3000 attempted bead types, a total of 2620 successful beads were generated (Data S1) for SNPs with more or less even distribution across the genome (Figure 3). The average distance between markers was 252.59 kbp, with a minimum distance of 15 bp and a maximum distance of 4848.33 kbp between a pair of SNPs in the centromeric region of chromosome Sb01. The number of markers per chromosome corresponds to the chromosome lengths; for example, the longest sorghum chromosome Sb02 was represented with the highest number of 338 SNPs per chromosome. The successful beads were generated in combination with SNPs from *Allium* species. During the design of the array, the sorghum and *Allium* SNPs were coselected to minimize interspecific cross-hybridization. Interrogation of the arrays and clustering was

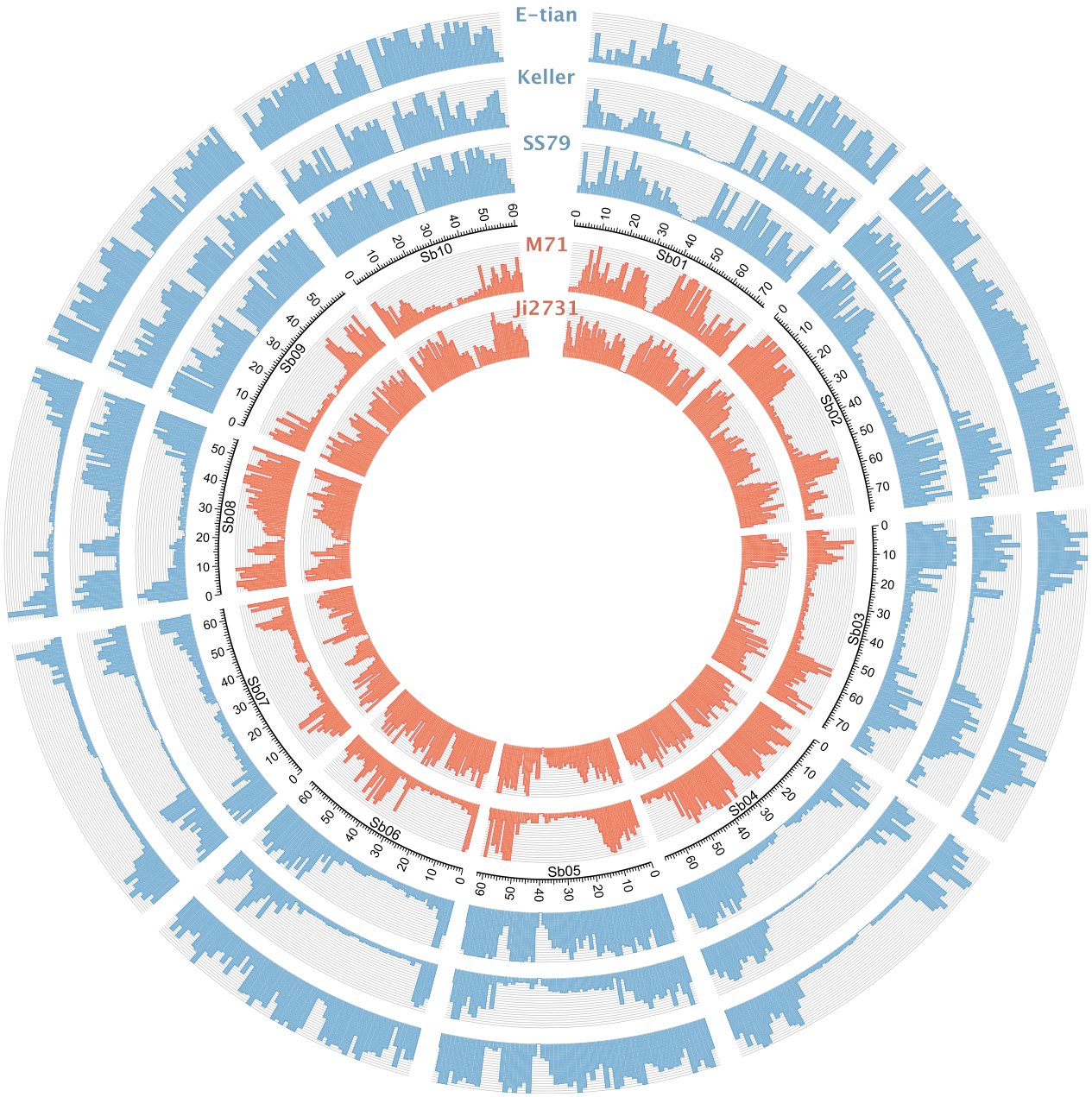


Figure 2 Overview of the genomic distribution of single-nucleotide polymorphisms discovered in five whole-genome sequences from grain (red histograms) and sweet (blue histograms) sorghum genotypes aligned to the grain sorghum reference sequence of BTx623. Histograms are smoothed at a resolution of 1 Mbp. The ten *Sorghum bicolor* chromosomes are drawn to scale in Mbp.

nevertheless performed separately with sorghum and *Allium* targets, and all *Allium* SNPs were zeroed during the clustering of the sorghum SNPs.

Clustering and call rates

From 3000 attempted bead types, a total of 2620 (87.3%) resulted in successful assays and up to 93.2% of these could be successfully called in the plant materials we screened. This conversion rate of 81.4% corresponds to the expected average design conversion rate of 80% for Illumina's Infinium assays. The observed call rates ranged from 77.8 to 93.2 in the different populations we screened, giving an average call rate in the 576 samples of 88.96%.

The average ADT score of all attempted sorghum SNP bead types, including failed bead types and SNPs with a call rate of zero, was 0.98. Hence, the ADT score was ruled out as a potential reason for SNP failure at any stage. The reproducibility error rates, calculated using the control sample (SS79) replicated in all plates, were well within the Illumina quality specification ($\leq$ 0.005) defined for validated human SNP Infinium genotyping.

SNP array characterization and comparison of SNP detection methods

A total of 760 (29%) of the SNPs in our array panel were derived from the bulked segregant sequencing, while the remaining 1860

Table 2 Chromosomal distribution of the 2620 iSelect array bead SNPs, which passed the Illumina quality tests, along with their projected positions in the sorghum reference genome sequence

Chromosome	Number of markers	Minimum distance (kbp) between markers	Mean distance (kbp) between markers	Maximum distance (kbp) between markers
Sb01	278	2.70	265.32	4848.33
Sb02	338	0.10	230.87	2178.98
Sb03	296	0.65	251.26	4549.23
Sb04	255	1.28	267.53	4553.42
Sb05	285	2.55	218.82	2758.09
Sb06	196	0.02	316.75	3289.38
Sb07	200	0.84	322.93	4091.32
Sb08	255	1.87	217.45	2725.55
Sb09	264	1.87	217.45	3327.44
Sb10	253	1.87	217.45	2725.55
Total	2620	0.02	252.59	4848.33

(71%) were genome-wide SNPs. From the latter SNPs, 76% (1415) were found to have a minor allele frequency of more than 1% in the screening panel, whereas only 56.3% (428) of the SNPs derived from the bulked sequencing had a minor allele frequency greater than 1%. The lower polymorphism in the bulk-derived SNPs can be explained by the lower diversity that was considered in selecting these SNPs, where only the polymorphism within SS79 × M71 RIL population in comparison with the reference genome was used rather than six available genomes used for the genome-wide selection of SNPs with a 50% allele frequency. Natural or artificial selection at a locus is known to cause reduced diversity and increased Linkage disequilibrium (LD) (Oraguzie *et al.*, 2007). Because most of the lines screened in this study are part of a breeding programme for Germany, these lines might have been indirectly selected for these genes, and as a result, the markers could not be polymorphic for these loci. One nice example for such reduced diversity due to selection in sorghum is the case of low diversity in sorghum conversion lines (long day insensitive lines) near the maturity locus on chromosome 6 (Morris *et al.*, 2013). Detailed validation of the failed and monomorphic markers using PCR-based methods or other SNP array technologies could give better insight into the complex segregation and inheritance of such SNPs, especially in the case of the SNPs derived from the bulk sequencing analysis.

Despite the reduction in polymorphism in the bulk-derived SNPs, these still showed a good conversion and quality rate. This high accuracy of the SNP detection presumably arose from the parallel validation of the SNPs detected in SS79 and M71 using two alternative SNP calling algorithms (CLC and SHOREmap). The use of genome-wide sequence data for SNP selection also enables an extremely stringent quality score filtering, because the number of SNPs is a virtually nonlimiting factor. Selection of markers developed from diverse lines and resources has been applied with success to develop SNP arrays for numerous species, including cattle, maize, chickpea and tomato (Ganal *et al.*, 2011; Hiermath *et al.*, 2012; Matukumalli *et al.*, 2009; Sim *et al.*, 2012). It is worth noting here that the array design scheme we followed gave a very high rate of successful and polymorphic markers; hence, the same scheme can readily be scaled up for the development of higher density arrays using the same SNP selection list. This opens the possibility for different scenarios, for example selection of SNP panels to target recombination-rich genome regions at higher density for genomic selection or genome-wide association studies.

Array polymorphism and its application for population classification

After automated clustering in the 564 genotypes (Table S1) using the default cluster file, only 127 (4.85%) of the 2620 scorable markers were found to have a call rate of 0 (PIC = 1). Visual screening of the image files revealed almost perfect clustering for all SNPs into the three clusters expected for a simple diploid organism like sorghum. Hence, we made no alterations to the cluster file except for removal of failed SNPs. A total of 1843 (70.34%) SNPs (MAF > 0.01) were polymorphic at the entire

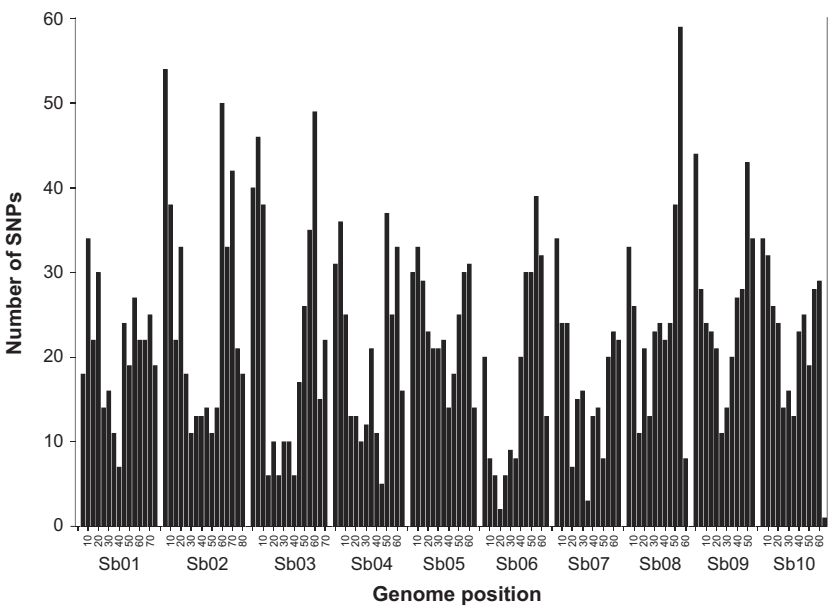


Figure 3 Genome distributions of the 3000 single-nucleotide polymorphisms (SNPs) selected for the array. The *x* axis indicates 5 Mbp intervals along the ten *Sorghum bicolor* chromosomes, while the *y* axis represents the frequency of selected SNPs within each 5 Mbp bin.

screening array level (Table 3). The average PIC value obtained within 1841 polymorphic markers in the diversity panel (n = 208) was 0.20, with a range from 0.0096 to 0.65. From the total of 1683 (64.2%) SNPs, which had a minor allele frequency (MAF) greater than 0.01 within the diversity panel (n = 208), 1574 (85.4%) markers had a minimum call rate of 80%. Detailed SNP array information on call rate and allele frequencies of SNP markers at a whole genotype set (n = 564), polymorphism in the mapping population (n = 92) and the diversity panels (n = 208) and its subpopulations (n = 154 and n = 54) are listed on the Supplementary Table S1.

Figure 4 shows a neighbour joining dendrogram constructed using 1843 SNPs with MAF > 0.01. The dendrogram clearly distinguishes the four mapping populations and the diversity panel. The latter showed the expected high diversity, described previously by Fiedler *et al.* (2012) using Diversity Array Technology (DArT) markers, while the clustering of the segregating populations reflects the genetic relationships among the parental lines with regard to the variation spanned by the diversity panel.

Population structure and linkage disequilibrium

A total of 1574 SNPs (with MAF > 0.01 and a minimum call rate of 80%) were used for population structure analysis on the diversity panel (n = 200) and in the parental lines from the mapping populations (n = 8). The ΔK derived from the structure analysis reveals the best cluster at K = 2, dividing the panel into two genetically different subpopulations with 154 and 54 genotypes, respectively (Figure S1). This supports the results of a previous analysis of 194 accessions from the diversity panel using DArT markers, which also revealed two subpopulations (Fiedler *et al.*, 2012). A net nucleotide distance of 0.1240 was revealed between the clusters, with expected heterozygosity of 0.08 within subpopulation/cluster 1 and 0.30 within subpopulation/cluster 2. The parental lines from the four mapping populations all grouped into the second subpopulation contributing to the higher diversity observed within the cluster.

Patterns of linkage disequilibrium were studied in the same set of genotypes (n = 208), using the same 1574 markers on the entire panel as well as within the two subpopulations separately using the corresponding markers with MAF > 0.01 and minimum call rate of 80% (Table 3). A mean r^2 value of 0.052 was calculated for the entire diversity set, while the smaller subpopulation 2 (n = 54) showed an intermediate mean r^2 value of 0.047 and the larger subpopulation had a mean r^2 value of only 0.034. However, the number of polymorphic loci between the 54 lines in the smaller subpopulation was higher than in both the larger subpopulation and the whole population.

The average LD decay (r^2 = 0.05) reflects the high level of recombination in sorghum, which was already reported by several authors (e.g. Bouchet *et al.*, 2012; Morris *et al.*, 2013). Nevertheless, a closer look at LD decay on a subpopulation level showed that the LD decay in subpopulation 2 (with only 54 lines) decays slightly faster than the larger subpopulation 1 and the entire population (Figure 5). Using the formula of Breseghello and Sorrells (2006), the critical r^2 value calculated for the whole population on chromosome Sb01 was 1.976. The LD starts to decay after just a few bp; however, the distance at which the nonlinear regression line intercepts the critical value is around 400 kb. This represents quite a large extent of LD compared with the recently published results of Morris *et al.* (2013), which estimated the LD in sorghum to decay at 75–150 kb depending on the genomic region. In that study, however, many more SNPs were analysed in a collection of over 950 worldwide sorghum lines representing several races and agro-climatic conditions. In comparison, most of the lines in our study (154/208) were in one subpopulation with lower level of diversity. Hence, the higher LD results from the lower diversity, as for example found by Lu *et al.* (2011) in temperate maize compared with tropical maize germplasm.

As reported previously in many plant and animal genomes, not only were differences in LD observed between subpopulations and chromosomes, but also across the length of chromosomes. This is explained by the higher recombination frequencies at the distal ends of the sorghum chromosomes (Mace and Jordan, 2011). The highest levels of LD were found within the heterochromatic region surrounding the centromere. Recombination suppression rates of up to 33% were reported by Kim *et al.* (2005) in sorghum heterochromatin. For high-resolution genome-wide association studies (GWAS), the high level of LD decay in gene-rich regions of sorghum diversity panels calls for the use of panels with hundreds of thousands of markers. Morris *et al.* (2013) achieved this marker density using a genotyping-by-sequencing (GBS) approach, whereas whole-genome resequencing as performed in this study is able to potentially reveal all genome-wide SNPs. Genomic skim sequencing may become a viable alternative for GBS as costs for next-generation sequencing continue to fall.

On the other hand, scalable high-density SNP arrays continue to be the method of choice for many applications where large numbers of specific loci need to be repeatedly assayed in large plant populations. This is particularly the case for research organizations and companies without the bioinformatics capacity to deal with GBS data. When genotyping for genomic selection during crop breeding, an increase in the quantity and density of SNPs with GBS does not necessarily increase the accuracy of selection or prediction models. It is often more important to increase the training population size, meaning that low-cost genotyping of large sample numbers for a fixed set of loci may be the priority. Recently, for example, a demonstration of the application of GBS for genomic selection in wheat showed that a

Table 3 Summary of polymorphism rates in the tested populations among the SNPs on the 3k SNP Infinium array

Minor allele frequency (MAF)	Minimum SNP call rate (%)	Entire screening panel (n = 564)	Complete diversity set (n = 208)	Diversity set subpop. 1 (n = 154)	Diversity set subpop. 2 (n = 54)
>0.01	–	1843 (71%)	1683 (64%)	1683 (64%)	1838 (70%)
	80	1738 (66%)	1574 (60%)	846 (32%)	1699 (65%)
>0.05	–	1795 (69%)	968 (37%)	903 (34%)	1510 (58%)
	80	1694 (65%)	882 (34%)	529 (20%)	1640 (63%)

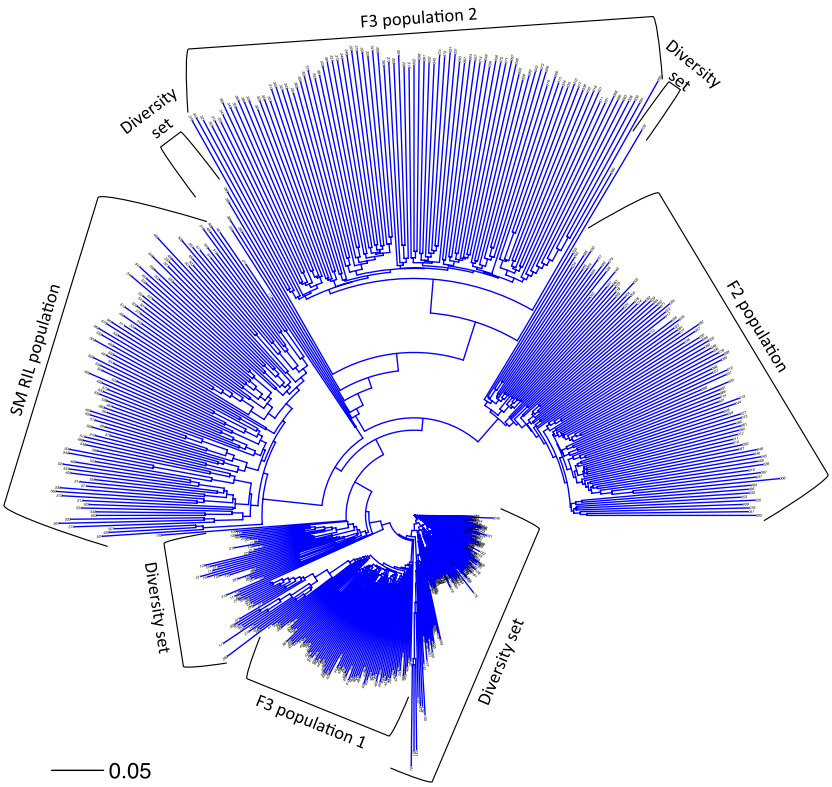


Figure 4 Neighbour joining dendrogram describing the genetic relationship among 564 sorghum genotypes using 1843 single-nucleotide polymorphism markers. The 564 lines consist of a diversity set, two F_3 populations, one F_2 population and a recombinant inbred line population as described in the Experimental procedures.

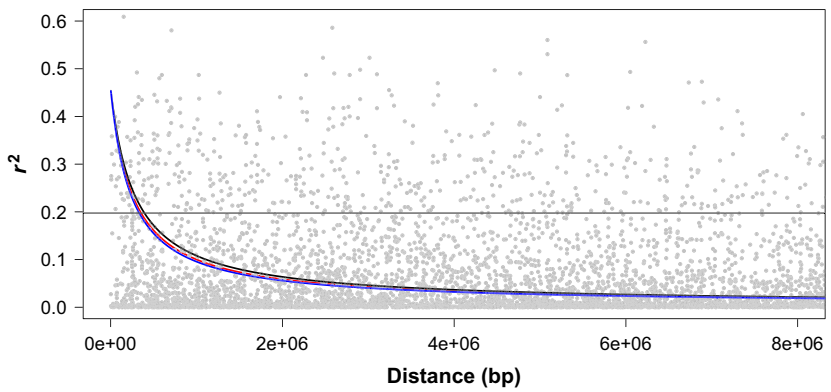


Figure 5 Decay in linkage disequilibrium (LD) in a sorghum diversity panel comprising 208 genotypes (black line) and in the subpopulations 1 (red line) and 2 (blue line), using 1574, 871 and 1693 markers, respectively, that showed minor allele frequency (MAF) > 0.01 and 80% minimal call rate (CR) in the respective genotypes. The line at r^2 = 1.97 shows the critical LD value for the total population.

reduction from 35 000 to 2000 GBS markers did not significantly reduce the prediction accuracy. On the other hand, the 2000 GBS markers did predict performance more accurately than the same number of DArT markers; this was believed to result from a clustering of many DArT markers to certain regions of the genome, whereas GBS markers are more randomly dispersed (Poland *et al.*, 2012). In this regard, our SNP array is also expected to be advantageous for genomic selection and predictive breeding applications, because the selection strategy resulted in an even distribution of SNPs throughout the entire genome (Figure 3).

The diversity panel tested in the present study was previously genotyped with DArT markers by Fiedler *et al.* (2012), who were only able to use 171 polymorphic loci for population structure and association mapping studies. The genome-wide SNPs we used increased the coverage with high-quality markers more than ninefold. Individual breeding programmes for many crop plants have a relatively low effective population size (N_e) due to long-term selection and limited access to adapted lines (e.g. with regard to day length dependency or low temperature tolerance in sorghum). As a result, many genomic selection studies reach a plateau of accuracy at only a few thousand markers. This means that the hundreds of thousands of SNPs that can potentially be obtained from GBS methods (e.g. Heffner *et al.*, 2011; Jannink *et al.*, 2010; Morris *et al.*, 2013; Poland *et al.*, 2012; Zhao *et al.*, 2012) may not necessarily increase prediction potential and may in fact unnecessarily increase the computational complexity of prediction models.

A high-density sorghum genetic map

The parental lines for the SS79 × M71 RIL population segregated for 1198 SNP markers. A very large proportion of these SNPs (1163, or 97.1%) were successfully placed onto a linkage map derived from 92 F_8 RILs. Regression ordering of the linkage groups resulted in an average marker density of around 1 marker per cM, demonstrating the usefulness of the array for development of dense linkage maps with sequence annotations to the sorghum genome sequence (Table 4 and Figure 6). The physical and linkage map orders of the markers corresponded in most cases (Figure 7); whereby, the conservation of the physical and genetic marker orders was greatly improved by linkage analysis using the maximum likelihood mapping (MLM) function of Joinmap 4 rather than regression mapping. Maximum likelihood mapping is reported to be extremely sensitive to missing data, however (Cheema and Dicks, 2009; Jansen *et al.*, 2001), causing inflated genetic distance estimates. This explains the exaggerated total length of the map we calculated by MLM, which covered more than 2161.14 cM compared with only 1068 for the map calculated by regression. On the other hand, the improved marker order achieved by the MLM function could be particularly clearly shown for chromosomes Sb01 and Sb03 (Figure 7). Similar physical and genetic mapping colinearity were reported in maize and tomato (Ganal *et al.*, 2011; Sim *et al.*, 2012). Only 17 markers (1.46%) were assigned to another chromosome than their projected positions. Six of these SNPs gave additional blast hits (e-10) to the same chromosome, suggesting possible duplications, while the remainder had multiple blast hits on several chromosomes.

Chromosomes Sb04 and Sb07 were each fragmented into three linkage groups; however, the marker orders within the linkage groups were colinear with the presumed physical positions. Similar fragmentation of linkage groups was reported previously during high-density genetic mapping (Hiermath *et al.*, 2012); however, this is generally not expected to impede accurate QTL mapping. Analyses of allele frequencies in the SM-RIL mapping population indicate a region of distorted allele frequencies on the long arm of chromosome Sb04, which probably led to the difficulties in linkage mapping in this region. The mapping parents also show low levels of polymorphism in a large block on chromosome Sb07, which was reported by Morris *et al.* (2013) to have low heterozygosity in sorghum.

Finally, our use of large numbers of more or less equidistant markers meant that we were able to accurately map local relationships between genetic and physical map distance. This is another important consideration when considering strategies for map-based cloning of target genes. As in maize, tomato and chickpea maps generated using high-density arrays (Ganal *et al.*, 2011; Hiermath *et al.*, 2012), many heterochromatic regions showed lower recombination frequencies compared with the euchromatic regions.

Conclusions

Availability of the reference sorghum genome sequence has paved the way for low-cost resequencing and identification of genome-wide SNPs that can potentially enhance genetic analysis and the application of molecular markers in sorghum genomics and breeding. Comparison of different SNP detection strategies revealed the feasibility of detecting high-quality, highly polymorphic SNPs even with low coverage sequencing. Furthermore, we demonstrated the enormous power of phenotypic pool sequencing for detection of trait-associated or QTL-linked SNP panels. Alignment to the genome sequence of a high-density genetic map, containing many equidistant SNPs, provided basic knowledge that will be useful to identify and characterize candidate regions for map-based cloning, using case-by-case experimental designs that reflect local LD levels. Analysis of 564 genetically diverse sorghum accessions including a diversity panel and different mapping populations revealed the usefulness of our SNP array for forward genetic analysis and genomic selection strategies in sorghum breeding populations.

Table 4 Summary of the SM-RIL *Sorghum bicolor* linkage map generated with the regression marker order function. A total of 35 additional polymorphic markers remained unlinked

Chromosome	Number of markers	Distance (cM) regression marker order	Marker interval		
			Min.	Max.	Mean
Sb01	122	77.973	0.00	3.67	0.64
Sb02	134	114.094	0.00	6.08	0.85
Sb03_a	4	0.000	0.00	0.00	0.00
Sb03_b	116	87.824	0.00	6.44	0.76
Sb04_a	22	36.632	0.00	10.09	1.67
Sb04_b	74	72.705	0.00	6.25	0.98
Sb04_c	15	4.307	0.00	1.43	0.29
Sb05	143	128.124	0.00	8.32	0.90
Sb06_a	4	8.101	1.57	4.28	2.03
Sb06_b	84	104.490	0.00	8.19	1.24
Sb07_a	19	42.729	0.261	7.233	2.25
Sb07_b	42	21.578	0.00	3.43	0.51
Sb07_c	13	30.022	0.212	14.036	2.31
Sb08	131	109.908	0.00	11.10	0.84
Sb09	114	111.440	0.00	7.90	0.98
Sb10	126	135.230	0.00	16.20	1.07
Total (whole-genome)	1163	1085.157	0.00	16.20	0.93

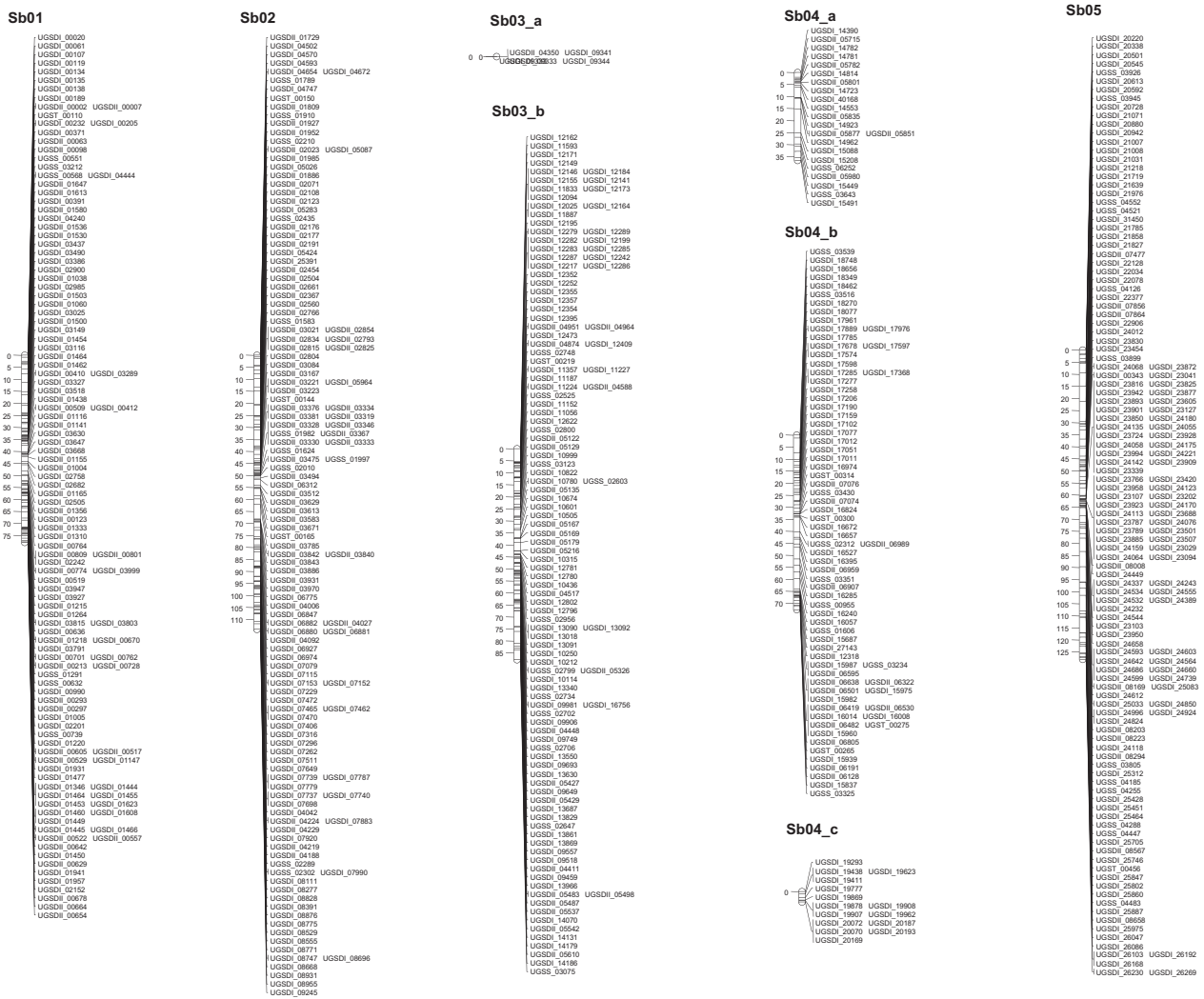


Figure 6 Genetic linkage map generated using single-nucleotide polymorphism array markers in 92 sorghum lines from the recombinant inbred line population SM (SS79 × M71). A total of 1163 markers were mapped into 16 linkage groups using regression for marker ordering. The average distance between markers is 0.98 cM, with a total linkage distance of 1098 cM.

Experimental procedures

Plant materials for SNP discovery

Raw sequence data from the resequencing of the three *S. bicolor* accessions described by Zheng *et al.* (2011) were made available prior to publication by the Chinese Academy of Sciences, Institute of Botany (Beijing, China). These included the sequences of the successful US sweet sorghum variety Keller, the Chinese sweet sorghum variety E-Tian and the Chinese grain sorghum variety Ji-2731. These three genotypes, which were resequenced by Zheng *et al.* (2011) to approximately 12× coverage with Illumina 100 bp paired-end sequencing, were complemented by whole-genome resequencing of the sweet sorghum line SS79 and the grain sorghum line M71, both from southern Africa. SS79 is an advanced sweet sorghum inbred line derived from a plant collected in a farmer's field in Limpopo, South Africa, while the grain sorghum inbred line M71 (Macia SA) originates from an ICRISAT breeding programme in Zimbabwe (Shiringani *et al.*, 2010). These two lines show considerable diversity with regard to plant height, photoperiodicity and flowering time, early-stage chilling tolerance, stem sugar and fibre, along with panicle and seed traits. Including the reference genome of the elite grain sorghum line Btx623 (Paterson, 2008; Paterson *et al.*, 2009), our SNP discovery panel therefore consisted of three grain and three sweet sorghum genomes from diverse genetic and geographical origins.

Phenotype-based SNP discovery

In addition to the new whole-genome sequences of SS79 and M71, 60 F_8 RILs derived from a cross between SS79 and M71 were skim-sequenced at approximately 0.8× coverage. These 60 RILs were chosen by phenotypic screening to segregate for tolerance to prolonged early-season chilling stress, a key trait for European energy sorghum breeding.

Whole-genome sequencing of SS79, M71 and 60 SS79 x M71 RILs

High-quality, low-plastid, genomic DNA was extracted from *in vitro*-grown roots of SS79, M71 and the 60 selected RILs using DNeasy® Plant Mini Kits (QIAGEN GmbH, QIAGEN Strasse 1,

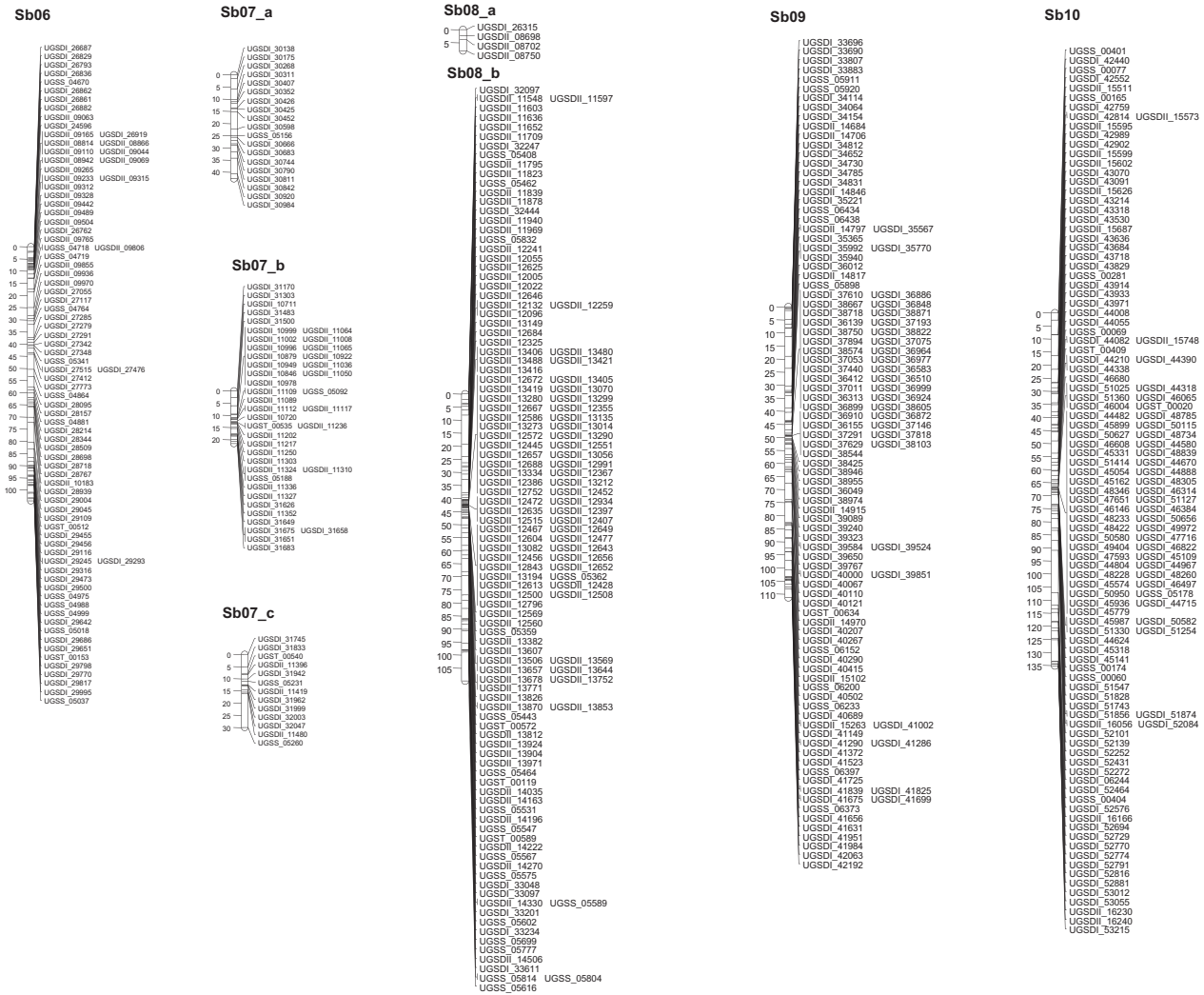


Figure 6 Continued.

40724 Hilden). DNA libraries were prepared using TruSeq Illumina® Paired-End DNA sample preparation kits (Illumina, Inc., San Diego, CA), which generate DNA fragments from 200 to 500 nt in length, followed by 95 bp paired-end sequencing.

SS79 and M71 were each sequenced on one lane of an Illumina Genome Analyzer IIx at Cologne Centre for Genomics (CCG), Cologne, Germany. Raw sequences were quality-filtered, trimmed and paired using the following parameters: Ambiguous = trim; Maximum number of nucleotides in reads = 1000; minimum number of nucleotides (nt) in reads = 15; Quality score = Illumina Pipeline 1.5 and later.

The phenotype-based sequencing of 60 RILs was undertaken at Beijing Genome Institute (BGI, Shenzhen, China) using random paired-end sequencing of nebulized and bar-coded fragments. The sequencing run was performed on a HiSeq 2000 sequencer. A total of 360 251 856 quality-trimmed reads were obtained with an average read length of 89.96 and an average distance between pairs of 470.87 nt (Table 1).

Sequence analysis and SNP detection

Paired-end reads with insert sizes ranging from 180 to 440 nt were filtered and paired if at least one of the sequences matched perfectly. The average insert length after mapping was 345 bp. The reads (average 93 bp) were mapped to the sorghum reference 79 accessed from Phytozome v9.0 (Goodstein *et al.*, 2012), using the CLC Genomics Workbench Version 4.9 (CLC Bio Aarhus, Denmark) set to the following mapping parameters: Similarity = 0.8, length fraction = 0.5, insertion cost = 3, deletion cost = 3, mismatch cost = 2, global alignment = yes. Inclusion of broken reads is not expected to change the accuracy of mapping (Li and Homer, 2010). The conflict resolution for consensus sequence formation was 'vote', that is, the more frequent nucleotide will be reported, however conflicts will be noted in downstream analyses such as SNP detection. Raw sequences obtained from Chinese Academy of Science were also filtered and mapped following the same pipeline. Due to the higher depth of coverage (12×), in this case, we used the 'join maps' function of the software CLC Genomics Workbench version 4.9 after mapping the reads from each genotype into three subgroups. CLC Genomics Workbench uses a neighbourhood quality score (NQS) method for SNP calling (Altshuler *et al.*, 2000). We applied the following parameters: Window length = 11 nt, maximum gap and mismatch = 2, central quality = 20 and average quality = 15. A SNP was called when it was detected in a

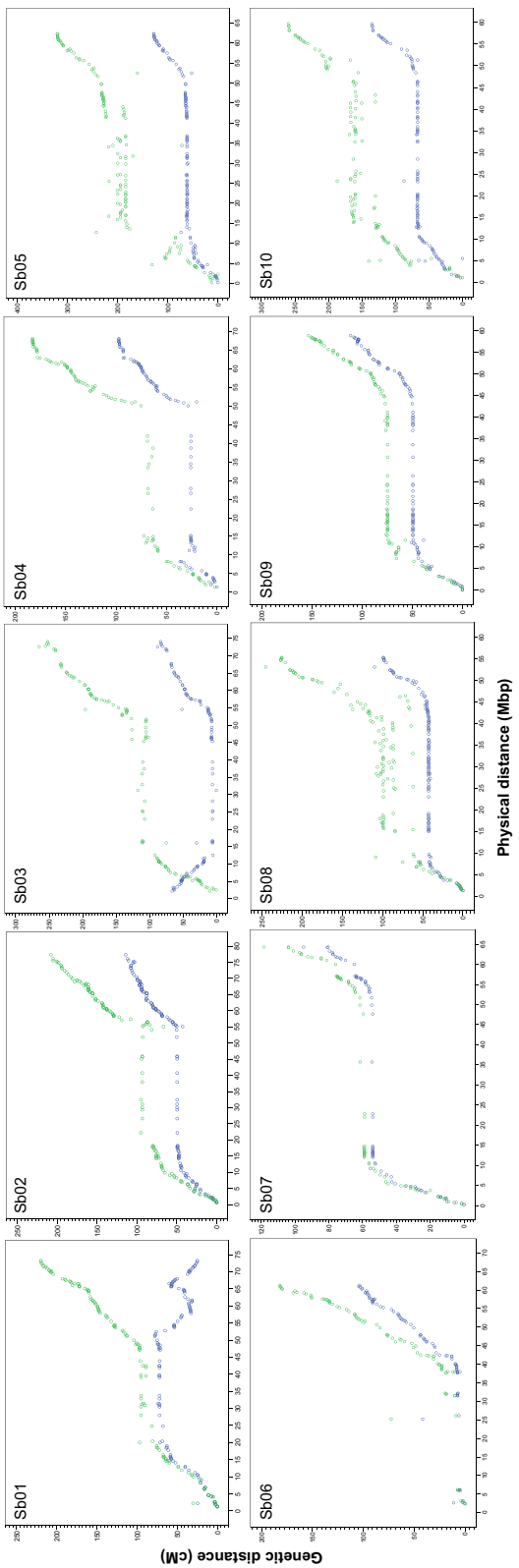


Figure 7 Colinearity between genetic and physical distance high-density single-nucleotide polymorphism (SNP) maps generated from the SM-RIL population (SS79 × M71). Blue circles indicate map positions calculated by a linear regression map order function, while the green circles show the map order results from maximum likelihood mapping. The latter resulted in a considerably improved fit of SNP marker orders to the expected positions within the sorghum genome.

minimum of four reads with at least 35% allele frequency and no more than two bases per locus. The CIRCOS circular genome presentation software (Krzywinski *et al.*, 2009) was used to plot the histogram of SNPs detected in five grain and sweet sorghum genotypes.

For technical validation, a parallel SNP detection was performed on the SS79 and M71 read data using the SHORE mapping and analysis pipeline for Illumina sequences onto reference sequences (Ossowski *et al.*, 2008). Only those SNPs that were detected using both CLC and SHORE mapping, and having a high genotype score of Q40 or greater, were used for downstream analysis and SNP selection.

SNP calling in phenotype-based RIL pools

Single-nucleotide polymorphism calling in the phenotype-based RIL pools was performed in two repetitions, once with highly stringent calling parameters and once with relaxed parameters. For the stringent calling, the minimum sequence coverage for the SNP detection in the two RIL pools was set to 10 reads, while a minimum SNP frequency of 99% was set for each group to ensure that only one allele per pool was permitted to be detected. Single-nucleotide polymorphism identification was subsequently performed once again using relaxed criteria, with a minimum coverage of four reads and a minimum SNP frequency of 2.4%, respectively. Single-nucleotide polymorphisms that were detected under stringent conditions at high frequencies in one group, and where the opposing allele was not detected under relaxed stringency in the second group, were considered to be unique to the pool in which they were found. Gene ontology (GO) term enrichment analysis of the genic SNPs was performed using the agriGO online analysis toolkit and database (Du *et al.* 2010).

SNP selection and array characterization

Candidate SNPs from the whole-genome data were filtered and selected by applying stringent criteria for genotype quality, allele frequency, the Illumina Assay Design Tool (ADT) score and the distribution throughout the genome. Besides the genome-wide SNPs, further SNPs were also selected in chromosome regions enriched with QTL for early-stage chilling tolerance, using the pooled sequence data. In this case, SNPs associated with the chilling-resistant genotype pool were preferentially selected based on proximity to potential candidate genes for abiotic stress tolerance and possibility of developing Infinium type II assays. The Infinium Type II assay requires only one probe to detect both alleles for the most common SNP types (A/G, A/C, T/G, T/C), in contrast to the less common type I SNPs (A/T and C/G), which require two probes per SNP.

Figure 1 gives an overview of the scheme used for the SNP selection from the genome-wide SNPs and the SNPs in the phenotypic pools. The genome-wide SNPs were selected for even distribution throughout the entire genome, with an intentional absence of bias towards gene-rich regions or coding sequences. These SNPs were complemented by trait-linked SNPs showing polymorphisms between the phenotypic pools for early stage chilling tolerance. Finally, the SNP density was increased in important QTL-rich regions identified from multitrait phenotyping of the SS79 × M71 RIL population (unpublished data).

SNP array screening and genotype scoring

High-quality DNA samples from the 564 genotypes of the screening population and the control genotype (SS79) were extracted using a CTAB extraction protocol modified from Doyle

and Doyle (1990). A total of 92 RILs from the cross SS79xM71 were used to validate the SNPs. Furthermore, a diversity panel comprising 200 *Sorghum* spp. Accessions, maintained by KWS Saat AG (Einbeck, Germany), and three genetically divergent segregating populations (one F_2, two F_3) were also genotyped using the array (Table 3).

Microarray-based DNA genotyping was performed by Service-XS B.V., Leiden, the Netherlands, using the custom-designed iSelect BeadChip (Illumina, Inc., San Diego, CA) interrogating 2620 Infinium Type II SNPs. For each sample, 4 μL genomic DNA at ~50 ng/μL was processed and hybridized to the BeadChips, according to the manufacturer's instructions. The BeadChip images were scanned on an Illumina iScan array reader, and the raw data were extracted into the Illumina GenomeStudio software (version 2011.1) using the default analysis settings. Genotyping analysis was performed using the Genotyping Module version 1.9.4 with the recommended default settings. A GenCall cut-off of 0.15 was applied, and clustering algorithm 2.0 was used. Clustering was performed using all samples and positive controls (with omission of the NTCs). Reproducibility error rates were calculated between the control sample replicates in each of the six screened plates.

The polymorphic information content (PIC) of the markers, which describes the measure of genetic diversity at a marker level dependent on the number of alleles and the frequency in a given population, was calculated in the 200 lines of the diversity panel plus the 8 mapping parents according to Anderson *et al.* (1993) using the following

$$PICi = 1 - \sum_j Pij^2 \quad (1)$$

where P*ij* is the frequency of the j_{th} allele for the i_{th} SNP marker.

Application of the array for population genetics analysis

Population structure was analysed using the programme STRUCTURE 2.3.1 (Pritchard *et al.*, 2000) using the 200 lines from the diversity collection plus the eight parents of the four mapping populations. The analysis was set at a burn-in period of 10 000 and 10 000 Markov chain Monte Carlo iterations, with 20 iterations to test *K* from 1 to 20 with 10 iterations for each *K* group. The optimum *K* value was calculated using the Δ*K* system according to Evanno *et al.* (2005) using the online tool STRUCTURE harvester (Earl and vonHoldt, 2012).

Linkage disequilibrium

Linkage disequilibrium analysis was performed using the software TASSEL 4 (Bradbury *et al.*, 2007) in two subsets of the diversity panel identified by STRUCTURE analysis as well as the entire diversity set ($n = 208$). The LD was calculated for all pairs of markers, and subsequent dissection of the LD estimates into chromosomes was then used to calculate LD (R^2) in 5 Mbp bins based on the physical positions of the markers. Decay in LD was calculated using the modified R code LDit.r (Ross-Ibarra Lab, University of Davis, CA, USA; see http://www.plantsciences.ucdavis.edu/faculty/ross-ibarra/code/files/LDit.html), which uses equation 1 from Remington *et al.* (2001) to estimate C employing nonlinear least squares and then plot the decay.

The critical LD value was calculated by square root transformation of the r^2 values of the unlinked chromosome LD values and calculating the parametric 95th percentile according to Breseghello and Sorrells (2006). Linkage disequilibrium above this critical value is considered linked, and the interception point where the nonlinear regression model line meets this critical line is assumed to represent the population LD decay point.

Linkage mapping

Genetic linkage mapping was performed using the SNP calls from the 92 F_8 RILs from the RIL population SS79 x M71 (SM-RIL), applying Haldane's mapping function in the software JoinMap®4 (Kyazma, Wageningen, the Netherlands). The markers were grouped into linkage groups at LOD values from 8 to 15. Markers were ordered alternatively using regression and maximum likelihood methods, and the fit of marker orders for each mapping algorithm was compared with the sorghum physical map in relation to the projected positions of the SNP markers.

Acknowledgements

This work was partially funded by grant number 0315421B from the German Federal Ministry of Education and Research (BMBF), with additional support from the Federal Ministry for Consumer Affairs, Nutrition and Agriculture (BMVEL), Grant 23/12-13C HN12, and DAAD. The authors thank Hai-Chun Jing, Institute of Botany, Chinese Academy of Science, Beijing, China, for the provision of prepublication raw genome sequence data, Korbinian Schneeberger, Max-Planck Institute for Plant Breeding Research, Cologne, Germany, for assistance with SNP calling by SHOREmap, Wilbert van Workum (Service XS, Leiden, Netherlands) for generation of the SNP array and Janine Altmüller (CLC Genomics, Cologne, Germany) for generation of the genomic resequencing data.

References

Altshuler, D., Pollara, V.J., Cowles, C.R., Van Etten, W.J., Baldwin, J., Linton, L. and Landers, E.S. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, **407**, 513–516.

Anderson, J.A., Churchill, G.A., Autrique, J.E., Tanksley, S.D. and Sorrells, M.E. (1993) Optimizing parental selection for genetic linkage maps. *Genome*, **36**, 181–186.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A. and Johnson, E.A. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376–e3376.

Batley, J. and Edwards, D. (2007) SNP applications in plants. In: *Association Mapping in Plants*, (Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N., eds), pp. 95–102. New York: Springer.

Bouchet, S., Pot, D., Deu, M., Rami, J.-F., Billot, C., Perrier, X., Rivalllan, R., Gardes, L., Xia, L., Wenzl, P., Kilian, A. and Glaszman, J.-C. (2012) Genetic structure, linkage disequilibrium and signature of selection in sorghum: Lessons from physically anchored DArT markers. *PLoS ONE*, **7**, e33470.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, **23**, 2633–2635.

Breseghello, F. and Sorrells, M.E. (2006) Mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. *Genetics*, **172**, 1165–1177.

Cheema, J. and Dicks, J. (2009) Computational approaches and software tools for genetic linkage map estimation in plants. *Brief. Bioinform.* **10**, 595–608.

Chia, J.-M. and Ware, D. (2011) Sequencing for the cream of the crop. *Nat. Biotechnol.* **29**, 138–139.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M. and Blaxter, M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **12**, 499–510.

Doyle, J.J. and Doyle, J.L. (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Du, Z., Zhou, X., Ling, Y., Zhang, Z. and Su, Z. (2010) agriGO: a GO analysis toolkit for the agricultural community. *Nucl. Acids Res.* **38**, W64–W70.

Earl, D.A. and vonHoldt, B.M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* **4**, 359–361.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S. and Mitchell, S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE*, **6**, e19379.

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

Fiedler, K., Bekele, W.A., Friedt, W., Snowdon, R., Stützel, H., Zacharias, A. and Uptmoor, R. (2012) Genetic dissection of the temperature dependent emergence processes in sorghum using a cumulative emergence model and stability parameters. *Theor. Appl. Genet.*, **125**, 1647–1661.

Ganal, M.W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E.S., Charcosset, A., Clarke, J.D., Graner, E.-M., Hansen, M., Joets, J., Le Paslier, M.-C., McMullen, M.D., Montalent, P., Rose, M., Schön, C.-C., Sun, Q., Walter, H., Martin, O.C. and Falque, M. (2011) A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE*, **6**, e28334.

Goodstein, D.M., Shu, S.Q., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D.S. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186.

Heffner, E.L., Jannink, J.-L., Iwata, H., Souza, E. and Sorrells, M.E. (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Sci.* **51**, 2597–2606.

Hiermath, P.J., Kumar, A., Penmetsa, R.V., Farmer, A., Schlueter, J.A., Chamarthi, S.K., Whaley, A.M., Carrasquilla-Garcia, N., Gaur, P.M., Upadhyaya, H.D., Kishor, P.K., Shah, T.M., Cook, D.R. and Varshney, R.K. (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* **10**, 716–732.

Hu, Z., Hua, W., Huang, S., Yang, H., Zhan, G., Wang, X., Liu, G. and Wang, H. (2012) Discovery of pod shatter-resistant associated SNPs by deep sequencing of a representative library followed by bulk segregant analysis in rapeseed. *PLoS ONE*, **7**, e34253.

Jannink, J.L., Lorenz, A.J. and Iwata, H. (2010) Genomic selection in plant breeding: from theory to practice. *Brief. Funct. Genomics*, **9**, 166–177.

Jansen, J., Jong, A.G. and Ooijen, J.W. (2001) Constructing dense genetic linkage maps. *Theor. Appl. Genet.* **102**, 1113–1122.

Kang, J. and Turano, F.J. (2003) The putative glutamate receptor 1.1 (AtGLR1.1) functions as a regulator of carbon and nitrogen metabolism in *Arabidopsis*. *Proc. Natl Acad. Sci. USA*, **100**, 6872–6877.

Kim, J.S., Islam-Faridi, M.N., Klein, P.E., Stelly, D.M., Price, H.J., Klein, R.R. and Mullet, J.E. (2005) Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. *Genetics*, **171**, 1963–1976.

Krzywinski, M., Schein, J.E., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J. and Marra, M.A. (2009) CIRCOS: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645.

Li, H. and Homer, N. (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinform.* **11**, 473–483.

Lu, Y., Shah, L.Y., Hao, Z., Taba, S., Zhang, S., Gao, S., Liu, J., Cao, M., Wang, J., Prakash, A.B., Rong, T. and Xu, Y.(2011) Comparative SNP and Haplotype Analysis Reveals a Higher Genetic Diversity and Rapider LD Decay in Tropical than Temperate Germplasm in Maize. *PLoS ONE*, **6**, e24861.

Mace, E.S. and Jordan, D.R. (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and gene-rich regions with significant implication for crop improvement. *Theor. Appl. Genet.* **123**, 169–191.

Mace, E.S. and Jordan, D.R. (2013) Supermodels: sorghum and maize provide mutual insight into the genetics of flowering time. *Theor. Appl. Genet.* **126**, 1377–1395.

Matukumalli, L.K., Lawley, C.T., Schnabel, R.D., Taylor, J.F., Allan, M.F., Heaton, M.P., O'Connell, J., Moore, S.S., Smith, T.P.L., Sonstegard, T.S. and Van Tassell, C.P. (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE*, **4**, e5350.

Metzker, M. (2010) Sequencing technologies – the next generation. *Nat. Rev. Genet.* **11**, 31–46.

Morris, G.P., Ramu, P., Deshpande, S.P., Hash, C.T., Shah, T., Upadhyaya, H.D., Riera-Lizarazu, O., Brown, P.J., Acharya, C.B., Mitchell, S.E., Harriman, J., Glaubitz, J.C., Buckler, E.S. and Kresovich, S. (2013) Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl Acad. Sci. USA*, **110**, 453–458.

Murray, S.C., Sharma, A., Rooney, W.L., Klein, P.E., Mullet, J.E., Mitchell, S.E. and Kresovich, S. (2008) Genetic improvement of sorghum as a biofuel feedstock: I QTL for stem sugar and grain non-structural carbohydrates. *Crop Sci.* **48**, 2165–2179.

Nelson, J.C., Wang, S., Wu, Y., Li, X., Antony, G., White, F.F. and Yu, J.M. (2011) Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. *BMC Genomics*, **12**, 352.

Oraguzie, N.C., Wilcox, P.L., Rikkerink, E.H.A. and Silva, H.N. (2007) Linkage disequilibrium. In: *Association Mapping in Plants* (Oraguzie, N.C., Rikkerink, E.H.A., Gardiner, S.E. and De Silva, H.N. eds), pp. 11–38. New York: Springer.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N. and Weigel, D. (2008) Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18**, 2024–2033.

Paterson, A.H. (2008) Genomics of sorghum. *Int. J. Plant Genomics*, **2008**, 362451.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schmutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R., Penning, B.W., Salamov, A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gingle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Rahman, M., Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J. and Rokhsar, D.S. (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature*, **457**, 551–556.

Poland, J., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., Dreisigacker, S., Crossa, J., Sánchez-Villeda, H., Sorrells, M. and Jannik, J. (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*, **5**, 103–113.

Pritchard, J.K., Stephens, M. and Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Remington, D.L., Thornsberry, J.M., Matsuoka, Y., Wilson, L.M., Whitt, S.R., Doebley, J., Kresovich, S., Goodman, M.M. and Buckler, E.S. (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl Acad. Sci. USA*, **98**, 11479–11484.

Riedelsheimer, C., Czedik-Eysenberg, A., Grieder, C., Lisec, J., Technow, F., Sulpice, R., Altmann, T., Stitt, M., Willmitzer, L. and Melchinger, A.E. (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nat. Genet.* **44**, 217–220.

Rooney, W.L., Blumenthal, J., Bean, B. and Mullet, J.E. (2007) Designing sorghum as a dedicated bioenergy feedstock. *Biofuel Bioprod, Bior*, **1**, 147–157.

Shiringani, A.L., Frisch, M. and Friedt, W. (2010) Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L Moench. *Theor. Appl. Genet.* **121**, 323–336.

Sim, S.-C., Durstewitz, G., Plieske, J., Wieseke, R., Ganal, M.W., Van Deynze, A., Hamilton, J.P., Buell, R., Causse, M., Wijeratne, S. and Francis, D.M. (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. *PLoS ONE*, **7**, e40563.

Washburn, J.D., Murray, S.C., Burson, B.L., Klein, B.R. and Jessup, R. (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 . *Mol. Breed.* **31**, 153–162.

Zhao, Y., Gowda, M., Liu, W., Würschum, T., Maurer, H.P., Longin, F.H., Ranc, N. and Reif, J.C. (2012) Accuracy of genomic selection in European maize elite breeding populations. *Theo. Appl. Genet.* **124**, 769–776.

Zheng, L.Y., Guo, X.S., He, B., Sun, L.J., Peng, Y., Dong, S.S., Liu, T.F., Jiang, S., Ramachandran, S., Liu, C.M. and Jing, H.C. (2011) Genome-wide patterns of genetic variation in sweet sorghum and grain sorghum (Sorghum bicolor). *Genome Biol.* **12**, R114.

Supporting information

Additional Supporting information may be found in the online version of this article:

Figure S1 (a) Delta *K* (ΔK) analysis plot to determine the most likely population substructure within a set of 200 diverse *Sorghum bicolor* accessions analysed for genetic diversity with the 3k SNP Infinium array described in this study. The clear peak for ΔK at $K = 2$ suggests the division of the population into two major subpopulations. (b) Graphical representation of population substructure within a set of 208 diverse *Sorghum bicolor* accessions analysed for genetic diversity with the 3k SNP Infinium array described in this study. Two clear subpopulations were detected.

Table S1 Summary of sorghum genotypes screened using the 3k SNP Infinium SNP array.

Data S1 Excel spreadsheet containing flanking sequence information, assay design scores, genome positions, minor allele frequencies and call rates for 3000 sorghum SNPs used to design the 3k SNP Infinium array described in this study.

4 Discussion

4.1 Tapping the phenotypic and genomic diversity of sorghum

Sorghum, the 5th most important cereal crop in the world, is the only domesticated crop species that originates from Africa among the top ten crop plants. It is a crop with wide agro-climatic adaptation. The domestication history, coupled with sympatric gene transfers with its wild relatives in all regions where sorghum is produced, make the species amenable to local adaptation and selection to fit different production methods. This has contributed to both phenotypic as well as genotypic variability of the crop (Mace et al. 2013). The high phenotypic variation of sorghum resulted in the classification of the species into races depending on the panicle structure and spike structure. The five different races of sorghum are all present in Africa, alongside its wild relatives (Harlan and De Wet 1972; Dillon et al. 2007). The broad end-use diversity of sorghum (e.g. food, feed, fodder, broom sticks, etc.) has contributed to the diversification into extremely variable morphotypes. These include short, grain sorghum types that can easily be harvested by combine harvesters, and sweet biomass types that can grow up to 5 m high and have sweet, juicy stems. Comparison of the parental lines of the mapping population used in the present study clearly demonstrates the high degree of variation in sorghum.

The sweet sorghum biomass parent SS79 was collected from South Africa and shows characteristics similar to Kaffir phenotypes (W. Wenzel, pers. comm.). Its plant height is on average 2.4 m at the time of harvest. It requires on average 112 days to flowering, has a high level of tillering, and achieves high biomass yields. In contrast, the grain sorghum parent M71, a Caudatum type derived from a mass selection program of the Southern African sorghum breeding program, is only 1.2 m tall and has a high grain yield, white seeds and earlier flowering (95 days) (Shiringani et al. 2010). The parental lines as well as the RILs derived from their cross segregate for several bioenergy traits (Shiringani and Friedt 2011). The parental lines also showed a contrasting phenotypic variation or segregation for tolerance to early-stage chilling (Bekele et al. 2013a) and drought stress. M71, which was released in more than five Southern African countries, was derived from the Macia breeding population from the International Crop Research Institute for the Arid and Semi Arid Tropics (ICRISAT). The drought and chilling survival phenotype of M71, together with the root phenotype observed in this work, are the results of long-term genetic improvement by ICRISAT and regional breeding organizations (Setimela et al. 1997). On the other hand, the high biomass and sugar content of the sweet sorghum land race SS79, originating from the

Limpopo region of South Africa, is a good example for the presence of high morphological and genetic variability in local sorghum land races.

A collection of sorghum lines owned by KWS Saat AG (Einbeck, Germany), derived from crosses among different races of *Sorghum bicolor* ssp. *bicolor* and *Sorghum bicolor* ssp. *drummondii* (Sudan grass), were used for genome-wide association studies (Fiedler et al. 2012). Sudan grass is a good source of drought and chilling tolerant genotypes. Sudan grass derives from a natural hybrid between *S. bicolor* L. (Moench) and the wild *S. bicolor* ssp. *arundinaceum* (Desv.). Hence it is highly diverse and harbours alleles that are essential for adaptation to marginal areas (de Wet 1978; Mace et al. 2013; Morris et al. 2013). Traditional practices of Ethiopian and other African farmers, who intentionally grow wild and weedy species in the surroundings of their sorghum fields so that allele exchange between wild-crop sorghum continues, have been reported by several authors (e.g. Teshome et al. 1999). Murray et al. (2011b) showed that there is a very high level of diversity in wild sorghum in Kenya, with wild genotypes showing a dual mating system depending on the environmental conditions. Under suboptimal conditions the wild sorghum plants would ensure reproduction by selfing, with a switch to outcrossing under optimum climatic conditions in order to circumvent the inbreeding depression that accumulated during the harsh periods (Murray et al. 2011a). A recent comparison of Ethiopian wild and weedy species with ICRISAT conserved wild species called for improved *in situ* conservation efforts, particularly in countries where wild and weedy sorghum populations are threatened by high population pressure, the introduction of new crops or elite lines that outcompete existing land races or change traditional farming practices that preserve diversity from wild and weedy morpho-types (Adunga et al. 2013). Global *ex situ* or genebank accession of sorghum are currently curated by several national and international organizations. The USA hosts one of the largest collections, followed by ICRISAT, which currently keeps more than 36 thousand cultivated and wild accessions collected from 91 countries (Upadhyaya et al. 2009; Kimber et al. 2013).

Wild and weedy species of sorghum are essential sources of variation for biotic and abiotic stress tolerance, and for new crop resources such as perennial sorghum. The latter can be propagated over winter via tubers, a cultivation option which could potentially revolutionize bioenergy sorghum production in temperate regions (Washburn et al. 2013). Wild relatives of cultivated sorghum, such as *S. halpense* and *S. propinquum*, are also of use for understanding weed biology and crop domestication (Paterson et al. 2009). For example, Lin et al. (2012) used *S. bicolor* ssp. *bicolo*r x *S bicolor* ssp. *virgantum* crosses to identify the seed shattering

gene, one of the most important genes in the domestication of crop plants. Sequencing of a BAC library from a seed shattering *S. propinquum* genotype enabled fine mapping of the region and identification of the causal gene, a YABBY transcription factor on chromosome 2. The rice and maize homologs of the identified gene were also within seed shattering loci, harbouring haplotypes that determined the degree of seed shattering in the respective species. This is an excellent example for the application of genomic and, new sequencing technologies to mine the phenotypic diversity of the genus *Sorghum* for better understanding of cereal domestication. Understanding the biology of the noxious weed *S. halpense* will definitely help weed scientists, crop scientists and ecologists to learn from the strategy of this species to establish itself faster and better than any crop plant. The 25 species in the genus *Sorghum* have an extremely wide geographic adaptation that can be tapped in breeding for biotic and abiotic constraints; these species, as shown above, are essential components of cereal evolution and domestication studies (Dillon et al. 2007).

The phenotypic classification of sorghum races was recently supported by molecular data, with a few minor discrepancies, including identification of the East African Kaufira as a new group that was phenotypically classified as Guinea or Kafir (Brown et al. 2011). Apart from classification, understanding the genomic diversity of sorghum has paramount implications in sorghum breeding and improvement (Dillon et al. 2007; Paterson et al. 2009; Morris et al. 2013; Bekele et al. 2013b; Mace et al. 2013). It is with this aim that the sorghum genome sequence consortium was established in November 2004 (Kresovich et al. 2005). A draft genome was available in 2007, and annotation of the sequence of the reference genome BTx623 was completed in 2009 (Paterson et al. 2009). After rice this was the second grass genome to be completely sequenced, using a whole genome shotgun sequencing approach (Paterson et al. 2009). The genome sequencing effort was made possible by the rapid technological advancement in genome sequencing and alignment technologies (Nordborg and Weigel 2008; Michael and Jackson 2013). Recent technological advancements in next-generation sequencing make it possible today to resequence whole sorghum genomes for only a few hundred dollars apiece.

Since sorghum is the first fully sequenced C4 grass, with a relatively small diploid genome, it is intensively used as a model for the closely related energy and food crops sugarcane, Miscanthus and maize (Paterson et al. 2009; Kimber et al. 2013). The availability of the sorghum genome sequence encouraged several applied and basic scientists to use the genome as a template for whole genome resequencing. This enabled patterns of genetic variation to be

identified at a sequence level that are indicators of evolutionary history of sorghum during its domestication and breeding over the past 8000 years. Table 4.1 shows the major sources of nucleotide diversity (Buckler and Thornsberry 2002), one of the most common types of genetic variation. Sorghum's parallel domestication in several locations, followed by disruptive selection processes and its self-pollination that led to many locally and marginally adapted genotypes, have contributed to a decrease of diversity on local level – at least in comparison to maize. However, recent studies on wild grass species reveal that increased genetic diversity is the key for adaptation to environmental variation, especially in the warmer and drier climates encountered by sorghum (Fitzgerald et al. 2011). In addition to high mutation rates, a prerequisite to adapt to new agro-ecological conditions, the introduction of wild relatives of sorghum from Africa to the other continents where the crop expanded has facilitated a constant two-way exchange of alleles between the wild relatives and the domesticated crop (Mace et al. 2013). Though sorghum is mainly self-pollinated, differences in the rate of cross-pollination exist depending on the panicle compactness and wind conditions. This, together with constant gene flow from the wild, has increased the amount of genetic variation in the domesticated sorghum species. The result is a very high rate of genomic diversity within sorghum, which prevails as an extremely high frequency of SNPs and structural variations (Mace et al. 2013; Bekele et al. 2013; Morris et al. 2013; Zheng et al. 2011; Paterson et al. 2009). In the present study, resequencing of a sweet and a grain sorghum genome identified more than 1.12 million SNPs between the two genotypes. Their alignment with the grain sorghum reference genome enabled identification of 1.18M SNPs (1 SNP/1089bp) in between the sweet sorghum SS79 genome and the reference, compared to 0.63M SNPs (1 SNP/1097bp) between the grain sorghum M71 and the reference genome. The ranking is in accordance with what other authors also reported. Since SS79 is a land race and M71 is derived from a breeding program for Southern Africa, higher diversity or SNPs in SS79 than M71 was inevitable. Even though M71 had a lower number of SNPs compared to the reference sequence, almost 20% of the SNPs in M71 were in genic regions compared to only 9% genic SNPs detected in SS79. This is much higher than the proportion of genic SNPs reported by Mace et al. (2013). On the other hand, the total number of SNPs they reported from 42 *S. bicolor spp.* accessions was more than 4.9M SNPs and the total number of genic SNPs was much higher. The high SNP frequencies in genic regions in M71 could be due to the targeted combination of alleles from at least three different sources reported to be involved in Macia's pedigree: (SDS 3220), [(GPR-148 x E35-1)-5-1] x CS 3541] (Setimela et al. 1997). According to Table 4.1, line selection reduced the genome wide diversity of each

contributing parent, followed by positive selection that fixed alleles (SNPs) in genic regions from each contributing genotype. This was followed by further fixation of alleles during selfing of M71 from Macia.

Table 4.1 Factors that impact nucleotide diversity. Source: Buckler and Thornsberry (2002)

Factor	Correlation with diversity	Scope
Mutation rate	Positive	Often whole genome
Population size	Positive	Whole genome
Outcrossing	Positive	Whole genome
Recombination	Positive	Whole genome
Positive-trait selection	Negative	Individual genes
Line selection	Negative	Whole genome
Diversifying selection	Positive	Individual genes
Balancing selection	Positive	Individual genes
Background selection	Negative	Individual genes or whole genome
Population structure	Mixed	Whole genome
Sequencing errors	Positive	Individual genes
PCR problems	Negative	Individual genes

In addition to SNP, structural rearrangements are also of interest to identify important copy number variations (CNV) and insertion deletion (InDel) polymorphisms that are responsible for the differentiation of the sweet and grain sorghum types (e.g. Zheng et al. 2011). Though the level of sequencing depth in the present study was not sufficient to identify structural variations at a high confidence, a total of 195,165 and 220,776 deletions were identified in SS79 and M71, respectively, in comparison to the reference genome. However, the sequencing depth of M71 was somewhat lower than in case of SS79, thus it is difficult to differentiate true deletions from regions with insufficient sequencing coverage. Sequencing at higher depth will enable such differences to be called at higher accuracy, making them potentially useful for genetic analysis and breeding. In addition, 114,772 and 104,363 putative sequence amplifications/duplications were identified in SS79 and M71 in comparison to the reference genome (unpublished data). Even though differences in copy number variation are expected, consensus mapping can help to validate suspected CNV and InDels. An expression diversity study conducted by Jiang et al. (2013) on the reference genotype BTx623 and the sequenced sweet sorghum line Keller showed that phenotypic differences were mainly due to

divergence in gene expression, resulting from polymorphisms in regulatory regions such as promoters or 3' and 5' UTR regions, rather than CNV or presence-absence variations (PAV). The presence of major structural genome differences, driven mainly by transposable elements, is known to influence many traits. Prominent examples include aluminium tolerance in maize along with the classical maize seed color variation mediated by Ds/Ac transposable elements (McClintock 1951). Recent mapping of CNV in the barley genome showed that there is a higher rate of CNV in wild genotypes than cultivated lines. Genes involved in cell death and disease resistance, such as the nucleotide-binding leucine rich repeats (NBS-LRR) proteins, were among the highly represented genes showing CNV (Muñoz-Amatirane et al. 2013). Variation for copy number of abiotic stress-related CBF genes in winter cereals was shown by Knox et al. (2010) to be involved in low temperature responses, while map-based cloning of an aluminium tolerance QTL in maize revealed that a higher copy number of the *MATE1* gene increases aluminium tolerance (Maron et al. 2013). One of the first plant height and maturity genes cloned in sorghum, *DWARF3* (*Dw3*), carries a mutation caused by a duplication in part of the gene (Multani et al. 2003).

4.2 Genomics-enabled genetic analysis and breeding in sorghum

The availability of a reference genome sequence for sorghum has paved the way for fast and cost-effective resequencing to detect sequence polymorphisms that can in turn be applied for breeding or genetic studies. In the work described in this thesis (Bekele et al. 2013b), the reference grain sorghum was used as a template to identify patterns of genetic variation within two grain sorghum and three sweet sorghum lines. Genome-wide markers were selected for the development of a SNP array, along with additional SNP markers derived from QTL bulk sequencing of 60 RILs selected based on their survival under prolonged cold conditions (Bekele et al. 2013b). Numerous other studies have used the reference genome sequence for genetic studies including SSR marker development (Li et al. 2009), map based cloning (Lin et al. 2012) and whole genome resequencing of wild, cultivated and weedy species (Nelson et al. 2011; Zheng et al. 2011; Morris et al. 2013; Mace et al. 2013, Thurber et al. 2013). The sorghum genome has also enhanced research and development in other complicated energy grass genomes such as sugarcane (Souza et al. 2011) and Miscanthus (Ma et al. 2012), close relatives of sorghum that are now benefiting from its considerably simpler and smaller genome (Paterson 2013).

Genomic tools are generated by different groups or consortia for several purposes, ranging from human health (e.g. International HapMap Project), animal breeding (e.g. Bovine

Genome Sequencing and Analysis Consortium) and for different crop plants genetic analysis and breeding (e.g. arrays such as the maize 50K SNP array; Ganal et al. 2011). Especially crop plant and animal SNP arrays are in high demand from research organizations and breeding companies (Riedelsheimer et al. 2012) for genomic selection (GS) and genomic prediction purposes. Success in this area has attracted the attention of many organizations involved in plant breeding, including Consultative Group on International Agricultural Research (CGIAR) organizations such as the international wheat and maize research institute (CIMMYT) (Crossa et al. 2013). The other interesting aspect of these technological advancements is that they are also being applied to orphan crop species like cassava, which have low commercial value but a high global importance for food self-sufficiency (Prochnik et al. 2012; Varshney et al. 2012).

Next generation sequencing technologies have enhanced marker discovery and in several applied and basic science disciplines covering the areas of evolution, genome dynamics and function. A major advantage is the possibility to obtain large numbers of genetic markers at relatively low cost and within a very short period of time. One application is the deployment of genotyping-by-sequencing (GBS) for biparental mapping populations, as has been demonstrated in Arabidopsis, rice and sorghum (Huang et al. 2009). Next-generation mutation identification using the short-read (SHORE) mapping technology and other related applications have been used to identify mutations in Arabidopsis (Schneeberger et al. 2009). Krothapalli et al. (2013) showed that whole genome sequencing could be used to identify a chemically mutagenised gene involved in rapid cyanide release in sorghum. Innovative mapping-by-sequencing approaches in rice were used to pinpoint seedling vigour and cold tolerance QTL in rice (Takagi et al. 2013; Yang et al. 2013). In the work described in this thesis (Bekele et al. 2013b), genome-wide resequencing data from 60 RILs was also used to identify trait-related markers and QTL-linked candidate genes. Another interesting application of GBS is in the area of population genomics, where for example the global sorghum core collection was genotyped using 265,000 GBS-derived markers that were used to describe the structuring of sorghum populations into geographic as well as race groups. Genome-wide association studies (GWAS) on this population located historically important alleles such as the maturity alleles that were used to adapt sorghum to long day conditions (Morris et al. 2013). These results also enabled a retrospective study of the US breeding programs, which introgressed maturity and plant height alleles to adapt exotic sorghum collections to temperate grain production systems requiring mechanical harvesters (Thurber et al. 2013).

The rapid development of DNA marker technologies in plant breeding has paved the way for the application of marker-assisted selection (MAS) for major traits and genes. However, most important agricultural traits show quantitative variation that is difficult to capture adequately using a few markers that target only a few QTL. Although biparental QTL mapping can identify linked markers for MAS (for example the QTL colocalization region on chromosomes Sb01 and Sb06 in the present work; Bekele et al. 2013a), however such markers normally explain only part of the phenotypic variance and their transferability to different populations is very low (Jannink et al. 2010). QTL detection by GWAS in diverse sets of genotypes (e.g. Fiedler et al. 2012) can be more accurate in detecting alleles that are useful in broader genetic backgrounds. However, GWAS is usually confounded by population structure and biased effect estimates that can not be used to predict genotype performance. A possible solution for the aforementioned problems is the application of GS. Results of a GS experiment in the SM-RIL mapping population are discussed in section 4.5.

The application of the sorghum SNP array (Bekele et al. 2013b) for fine mapping and high density GWAS is shown on Figure 4.1. GWAS with 1494 markers on the SNP chip enabled identification of 12 SNPs associated with seedling emergence (P=0.05), confirming and refining most genome positions that were previously mapped by Fiedler et al. (2012) at low resolution with only 171 DArT markers. All chromosomes except chromosome Sb07 had at least one marker significantly associated with cold emergence, while the associations for emergence at optimum temperature were all below the threshold value. On the other hand, the significant association of marker UGSS_00632 with cold emergence on chromosome Sb01 (Figure 4.1) at position 14,709,234 bp corresponds to a projected QTL for controlled environment and field emergence (Bekele et al. 2013a) at the same position, and fine mapping within the 92 lines also identified QTL in closer regions (Sb01 around 17Mb) (Appendix III, Figure 1), and close to a previous low-resolution marker trait association (Fiedler et al. 2012). This marker, developed from the QTL bulk sequencing study, was unique to the pool of 30 chilling-susceptible RILs. Even though the marker itself was not inside a gene, the two flanking genes Sb01g015200 (Chromatin-associated protein Dek) and Sb01g015210, (similar to JmjC domain containing protein) are predicted to have a role in chromatin organization and epigenetic control of plant growth and development. The Arabidopsis orthologs identified by BLASTP analysis of Sb01g015200 included AT3G48710 (5e-35), a DEK domain-containing chromatin associated protein with roles in chromatin organisation and response to vernalization (Heyndrickx and Vandepoele 2012). Sb01g015200 has a BLASTP hit to the Arabidopsis gene AT3G07610, which encodes a protein with histone H3mK9 demethylation

activity that plays a crucial role in regulation of chromatin remodelling and methylation of genic and repeat elements (Saze et al. 2008). The co-localisation of genes involved in higher level genome organization, vernalization, gene expression and developmental regulation with QTL for chilling-related traits including field emergence, makes this an interesting region for future detailed studies. Additional potential QTL regions on chromosome Sb02, Sb06 and Sb09 also contain highly interesting candidate genes. Fine mapping using comparative QTL mapping, high-resolution GWAS and synteny analysis is in progress.

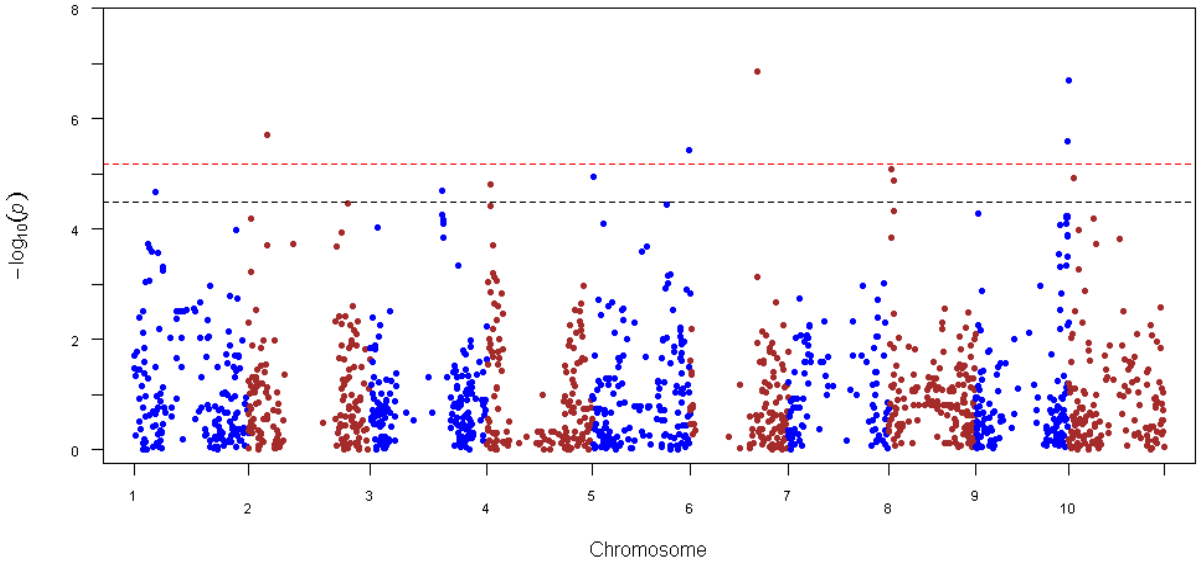


Figure 4.1 Marker trait associations for emergence under cold temperature regime (12.3°C). Phenotype data is from Fiedler et al. (2012) and 1494 MAF>0.01 SNP markers and population structure were described in Bekele et al. (2013b). The red (5.17) and black (4.48) broken arrows show the P=0.01and P=0.05 Bonferroni marker correction thresholds for the number of markers stated.

4.3 Genetic analysis of abiotic stress tolerance: The phenotyping bottleneck

As sessile organisms, plants are forced to devise abiotic stress response mechanisms to cope with environmental changes. Abiotic stress can be defined as an environmental disruption that significantly deviates from the optimal physico-chemical environment (light, water, temperature) of the plant and hence disturbs its physiological processes. Abiotic environmental factors determine the geographic distribution and the biological potential of a species. Abiotic stresses such as drought stress are estimated to cause up to 50% global yield reduction (Boyer 1982; Qin et al. 2011).

However, most important agricultural and fitness-related traits, including phenotypic plasticity and metabolic responses to abiotic stress, have a complex inheritance (Holland 2007). One of the bottlenecks for understanding the abiotic stress response is the difficulty to accurately quantify the morpho-physiological changes associated with abiotic stress response.

Technological improvements and innovative phenotyping methods are overcoming this bottleneck and morpho-physiological components of abiotic stress response are increasingly becoming quantifiable (Roy et al. 2011). The developments and wide application of molecular biological methods in plants has contributed its share to enhance our understanding of the molecular mechanisms of abiotic stress perception, signalling and reactions.

Abiotic stress analysis requires a good characterization of the stress in terms of severity, duration, sequence and combination of stresses the plant experiences in the field or in a controlled environment (Gaspar et al. 2002). It is also quite important to specifically describe the relevant plant characteristics, such as the organ, the developmental stage and the order and degree of stress on the plant. For example, a severe drought at seedling stages or flowering will have a different effect on yield and the plant phenology. Measurements on the fitness of the crop, its ability to reproduce or the amount of seeds it produces can give us the idea of how good the plant tolerated the stress and sustained its physiological processes. These responses are usually compared with the physiological or biological potential of the crop under optimal conditions. The reduction in fitness, or conversely the ability of the plant to maintain its growth, can be explained by the following two general mechanisms: (1) survival under extreme stress by reducing growth, and (2) maintenance of growth and development under mild stress conditions (Claeys and Inzè 2013).

The genetics of yield stability across different locations or environments are still not fully understood. Plant response to different environmental conditions has been studied from a whole-plant or ecological level down to cellular or metabolic changes that lead to tolerance/sensitivity or ultimate death of the plant or crop in question. Responses to drought, chilling and salinity stress have considerable overlap; hence these are jointly termed as osmotic stresses. Osmotic stress can be caused by lack of available water due to drought or salinity, but it can also be caused by the lack of transpiration reducing the ability of the plant to take up available water, especially under root chilling conditions (Aroca et al. 2001). There are several morphological and physiological processes that helped grasses adapt to almost all vegetated ecosystems. Grasses are known to harbour numerous genes that can help to adapt to chilling and even freezing conditions, as in the case of winter cereals. Chilling stress causes morphological, phenological and physiological changes that are perceived and regulated by complex molecular processes. As in any other stress, the mechanisms that are put in place for survival, growth or development depend on the intensity and duration of the stress and the genetics of the plant.

Sweet grasses are considered sensitive to chilling conditions, due to their C4 metabolism, however there are also tolerant species such as Miscanthus that are capable of producing high biomass in temperate regions. Recent phylogenetic analysis on 1230 C4 grasses showed that C4 photosynthesis is an adaptation to open fields and drier environments rather than *per se* warm temperature adaptation (Edwards and Smith 2010). Another study on evolutionary mutation rates in low temperature-responsive genes showed that five winter cereals/grasses in the Pooiedeae subfamily (*Brachypodium*, wheat, barley, *Lolium*, *Festuca*) exhibit a high rate of nonsynonymous mutations compared to the warm climate-adapted species (sorghum, maize and rice) (Vigeland et al. 2013). Though freezing tolerance is absent, there is definitely variability for cold tolerance in sorghum, especially in lines which are adapted to subtropical highland regions of Africa and the Middle East (Brar et al. 1992; Franks et al. 2006; Knoll et al. 2008, Payne et al. 2003).

Germination and emergence processes studied in a diversity panel and a mapping population showed that there is ample variability in the speed, uniformity and rate of emergence of sorghum under optimum (25/20°C) and chilling stress conditions (13°C) (Bekele et al. 2013a; Fiedler et al. 2012). Other seedling development processes were also shown to be retarded. In the absence of adequate resistance, prolonged chilling stress (13°C) was found to cause wilting and death of the seedling, however a huge variability was found for emergence, seedling shoot and root development and survival under chilling conditions. Simulation of early spring climatic conditions and dissection of the traits under controlled conditions helped to quantify the degree to which growth and development was reduced; this data could be used for identification of important main-effect QTL and QTL colocalization hubs. Interesting correlations of root developmental traits to survival and emergence under chilling conditions demonstrated the importance of root system architecture and conductance in early stage seedling development under chilling stress conditions in sorghum. Heterosis for sorghum biomass and seedling development, even under chilling stress conditions, has been reported in sorghum (Blum 2013). There are many researchers who think heterosis or hybrid vigour is caused by the robust root development of sorghum. The reports on the presence of heterosis for seedling vigour under chilling, and its impact on root development, show the importance of detailed genetic and physiological studies (Blum 2013). However, the *in situ* characterization of root system architecture and root growth dynamics remains a hard-to-overcome hurdle in the assessment of genomic effects on whole-plant fitness for survival and performance in changing environmental conditions (Dhont et al. 2013).

Chilling stress or nitrogen starvation leads to sink limitation, the condition where the plant can not produce enough sink to take up the carbohydrate fixed by the plant (Nunes et al. 2013). Development processes are known to be more sensitive to water limitation than photosynthesis, this could be the reason why fewer correlations and colocalizations were observed of chlorophyll content with other development related traits than between other growth related traits such as emergence and root length. Similar separate genetic control was reported by Hund et al. (2004) in maize. On the other hand, photosynthetic efficiency is also reduced during chilling stress, due to reduced chlorophyll fluorescence efficiency and chlorophyll content decline as in drought conditions (Hund et al. 2004; Leipner et al. 1997). One of the QTL for chlorophyll content under chilling stress conditions was colocalized with a root development QTL that is closely linked to the stay-green locus Stg2. The Stg2 locus has been shown to control leaf senescence during drought stress and is known to correlate with root angle, nitrogen metabolism and hormone balance (Harris et al. 2007, Mace et al. 2012).

The availability of a sorghum consensus map linked to the reference genome is a great advantage for sorghum geneticists and breeders (Mace et al. 2010). The QTL from this study (Bekele et al. 2013a) were projected onto the consensus map and compared with other cold tolerance studies in sorghum, rice and maize (Emma Mace, DAFFQ, Queensland, Australia, personal communication). From a total of 101 loci reported, 66 could be projected (Appendix III, Figure 2). The figure shows an example of a highly interesting region on chromosome Sb02, which was reported by several authors to include a major QTL for tolerance to low temperature in rice (Li et al. 2013). Despite high variation in the number of genotypes, the experimental setup and the genetic background of the mapping populations used, the possibility to detect similar regions over distantly related species is highly encouraging. This example also shows the power of combining controlled experiments with field phenotyping for genetic analysis of abiotic stress tolerance.

Field phenotyping and genetic analysis for complex traits such as yield is a demanding process that benefits from the presence of high density linkage maps and efficient controlled phenotyping technologies. Semi-controlled, high-throughput and high-dimensional phenotyping for traits like plant height that show high correlations between field environments and controlled systems can be employed for high-through put screening of breeding populations. For example, Figure 4.2 shows a major QTL for plant height on Sb07 that was previously mapped at seedling stage in rainout shelter experiments (data not shown)

and field experiments (Shiringani et al. 2010). This enabled an improvement of the QTL mapping from the field experiment using the high density map. The QTL for plant height under stress conditions explains about 80% of the phenotypic variance (Appendix III, Table1). Although there are some other main effects and epistatic QTL detected, the one on chromosome Sb07 explains 46.8% of the phenotypic variance and mapped close to the *Dw3* gene position. The QTL support interval of the QTL, from 56.2Mb to 57.2Mb, is only 1.6Mb from the causal mutation and the associated SNP detected in US lines genotyped by GBS (Thurber et al. 2013). Zhou et al. (2012) also used GBS, followed by bin mapping and QTL analysis for long and short day conditions, to accurately map the mutations *Ma1 and Dw3. Ma1* was mapped at 700 kb distance, while *Dw3* could be accurately mapped using 244 RILs and around 3000 bin markers. Considering the fact that the present example was from field experiments in a single year with only 92 RILs, it shows the very high power of QTL mapping using a high-density map, given the right combination of parental lines and a suitably segregating population. Thurber et al. (2013) recently took a retrospective look at the effect of the maturity and plant height introgressions into exotic sorghum lines, suggesting the presence of an additional plant height allele on chromosome Sb06 and describing the difficulty of mapping this locus. The authors recommended targeted introgression of alleles into a low-polymorphic haplotype of the chromosome Sb06 adaptation region in order to benefit from diverse alleles in this region. However, they also discussed the fact that the association for plant height at *Dw3* was not mapped directly at the gene position, suggesting a duplication which complicates mapping.

Further traits are of interest because of their relevance to sugar yield (Brix) and chlorophyll content at the 8th leaf stage, an indirect indicator of stay-green characteristics associated with drought tolerance mechanisms in varieties of sweet sorghum. The epistatic QTL cluster that explains a large proportion of the phenotypic variance for chlorophyll content on chromosome Sb03 is close to known stay green loci (Harris et al. 2007).

4.4 Genomics-assisted adaptation of sorghum to central European conditions

Sorghum is a very good alternative to maize for bioenergy production. The high biomass yield and its hardy nature have attracted the attention of breeders and scientists in Germany and other central European countries (Fiedler et al. 2012; Schittenhelm and Schroetter 2013). Sorghum was introduced to Europe a long time ago: Records showing the use of broom corn (sorghum) date back to 1204 in Italy (Dahlberg et al. 2011); however the lack of breeding to adapt the crop to European conditions has effectivley limited the production, especially in

central and northern Europe, to a very low extent. On the other hand, however, as maize production in Eastern and Central Europe expands there is an increasing threat from important pests, particularly the western rootworm *Diabrotica virgifera*. A major advantage of sorghum over maize production in Europe is the complete lack of maize western roomworm infestations on sorghum. Furthermore, Schittenhelm and Schroetter (2013) recently showed that sweet sorghum and Sudan grass hybrids are superior to maize under severe drought conditions in Germany and France rain out shelter experiments. This was attributed to the higher root biomass and deeper soil penetration of sorghum roots compared to maize.

Adaptation of a crop to a new agricultural system depends on an ability to integrate it into existing production and farming systems. The same agricultural equipments used for maize can handle sorghum production; hence it can be easily integrated into current maize-based cropping systems and can readily serve as a direct substitute for animal feed and biogas production. However, a new crop that is competing with a well-established crop has to meet high expectations set by the established system. Adaptation programs can benefit from any innovation that enhances breeding. For example, the application of modern genomics tools can potentially enhance efforts to produce lines or hybrids that are well-adapted and high yielding. Matching the high biomass yields and adaptation that were achieved in maize over 50 years of hybrid breeding programs represents a huge task for sorghum breeders and scientists, however the availability of sorghum genome resources today provides a completely different basis to repeat the success story of maize in a much shorter timeframe.

One of the advantages of temperate maize production is a relatively early sowing time that enables high biomass accumulation during the limited growing season in Northern Europe. As was the case when maize was introduced to temperate Europe, however, sorghum is sensitive to early stage chilling stress. Improvement is therefore needed to ensure emergence under low spring soil temperatures and achieve the high biomass and dry matter content standards set by energy maize production in Europe. Sorghum is currently sown from mid to end of May, whereas maize sowing takes place from mid-April until the end of June in case of double cropping systems, depending on the maturity group of the variety. In fact sorghum can also be sown late as a second crop, since it has a fast growth especially in the middle of the summer when temperatures are high enough for sorghum development. Based on these production scenarios the main current breeding goals in sorghum are early stage chilling tolerance, high yield, early flowering, lodging resistance and high dry matter content (> 27%) at harvest (KWS Saat AG).

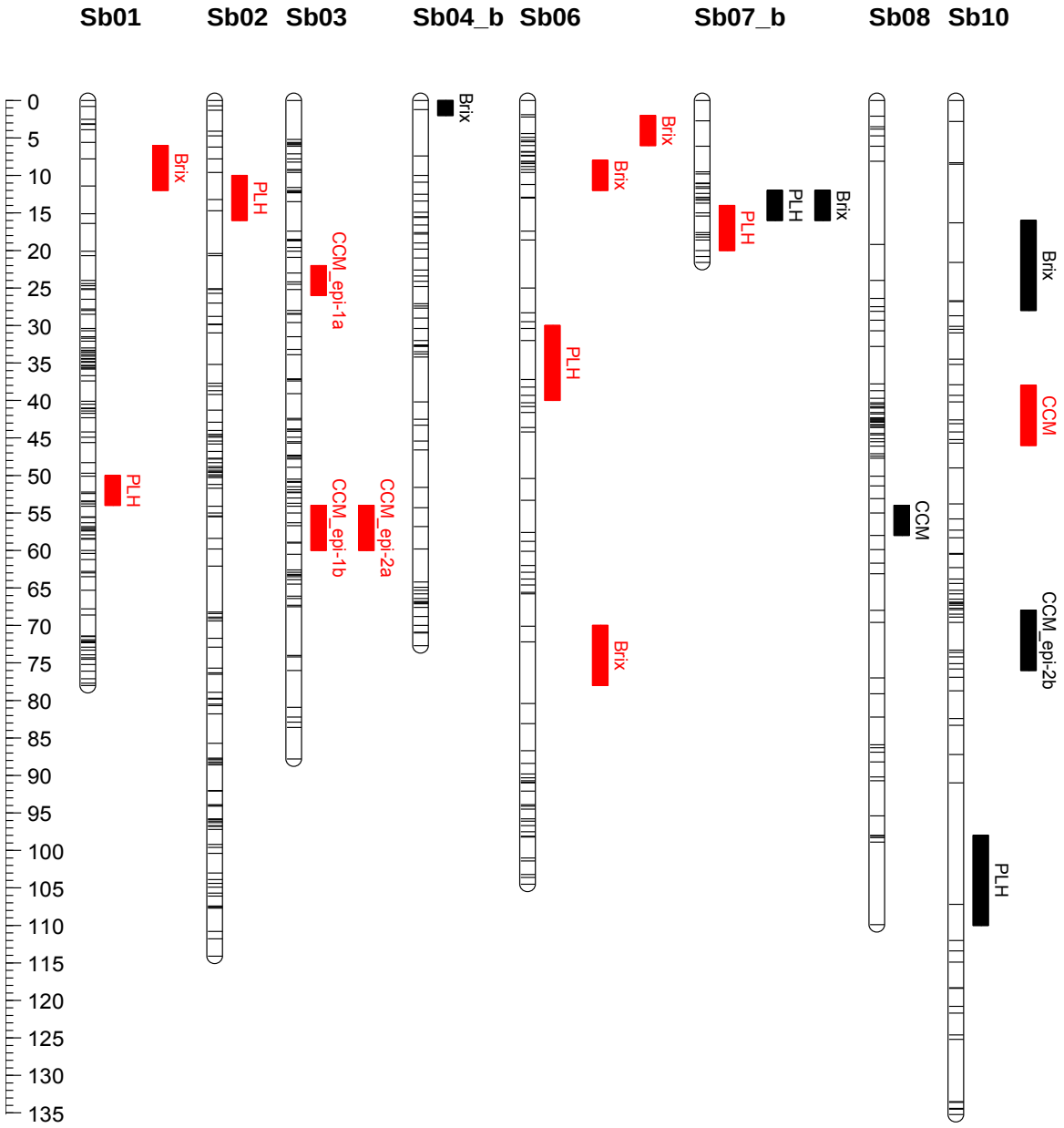


Figure 4.2 Quantitative trait locus (QTL) mapping using 1163 linked markers with trait data from a field drought experiment conducted at Gross Gerau, Germany, in the 2009 cropping season. The red (non-irrigated) and black (irrigated twice during the growing season) bars show the 1-LOD support interval of the detected QTL for different traits related to stress responses. Details of the field trials are given in the supplementary materials and methods (Appendix III), while the details of the QTL results are listed in Appendix III, Table 1. The traits considered are plant height at maturity (PLH), chlorophyll content measurement (CCM), and reduced sugar content (Brix). QTL with epistatic action are marked by the abbreviation "epi".

In the work described here, two traits of primary importance for sorghum adaptation and biomass production in Germany, i.e. emergence and plant height, are shown as examples for the application of genomics tools to enhance breeding under suboptimal conditions. Chapter 2 of this thesis showed the complexity of seedling emergence and seedling development under chilling conditions (Bekele et al. 2013a).

The genetic markers identified by association and biparental QTL mapping can be used in marker-assisted selection to enhance adaptation breeding for sorghum. Marker-assisted selection for cold tolerance was successfully undertaken in sorghum by Knoll and Ejeta (2008). However, the pyramiding of large numbers of significant but minor-effect QTL-linked markers, and the dynamic landscape of epistatic interactions depending strongly on the genetic background, triggered the need to instead use genome-wide marker systems.

One alternative tool that is possibly more effective for breeding of complex traits such as emergence under chilling conditions is whole-genome prediction. Considering the even distribution of our markers and a very high LD in bi-parental mapping populations (Bekele et al. 2013b), the SNP array was used for a proof-of-concept study of GS/prediction in sorghum. GS is a method where a training population is genotyped and phenotyped to develop a statistical model that gives a weight to each genome-wide marker depending on their predicted contribution to a given phenotype. This model is later used to predict the performance of candidate genotype or genomic estimate breeding value (GEBV) using only genotype data. Animal breeders first developed GS methods when genome-wide SNP screening platforms became affordable. They are now being used by several plant breeding groups to predict yield and other complex traits (Jannink et al. 2013). The two major criteria for a successful application of GS in plant breeding are the size of the training population and the marker density and distribution, together with the right statistical model for estimation of the phenotypic values from genotype data. Affordable high-throughput genotyping has increased the possibilities for application of GS; however recent studies showed that there is also a threshold where the accuracy would not significantly increase with further increases in the number of markers. This is especially the case in breeding materials with small effective population size and thus relatively conserved linkage disequilibrium. For example, Poland et al. (2012) recently showed that 2,000 markers could have predicted the phenotype in a wheat breeding program as accurately as 35,000 markers.

4.5 Proof-of-concept for genomic selection (GS) in sorghum

To demonstrate the effectiveness of GS for chilling tolerance traits in sorghum, the widely used ridge-regression best-unbiased linear prediction (rr-BLUP) model for genomic prediction was applied to the biparental SM-RIL population using the same markers already used for mapping and QTL analysis. As seen in Figure 4.3, even in a small population, GS can have reasonably high prediction accuracy for adaptation traits, in this case for emergence under cold/optimal conditions and biomass yield (plant height) in drought-stressed/irrigated

field trials, respectively. The results showed that it is possible to predict the performance for both of these traits with moderate to good accuracy even with relatively small training populations. Plant height under irrigated conditions and emergence under chilling stress showed mean accuracies in 500 iterations of up to 0.56 and 0.46, respectively. This is within the range of several other studies in different crops (Heslot et al. 2012). Several authors (e.g. Riedelsheimer et al. 2013; Heffner et al. 2011) have reported that increasing the size of the training population increased the accuracy of prediction. In the examples shown here, however, the improvement in prediction accuracy reached a plateau at a relatively low number of genotypes, particularly in case of emergence (Figure 4.3).

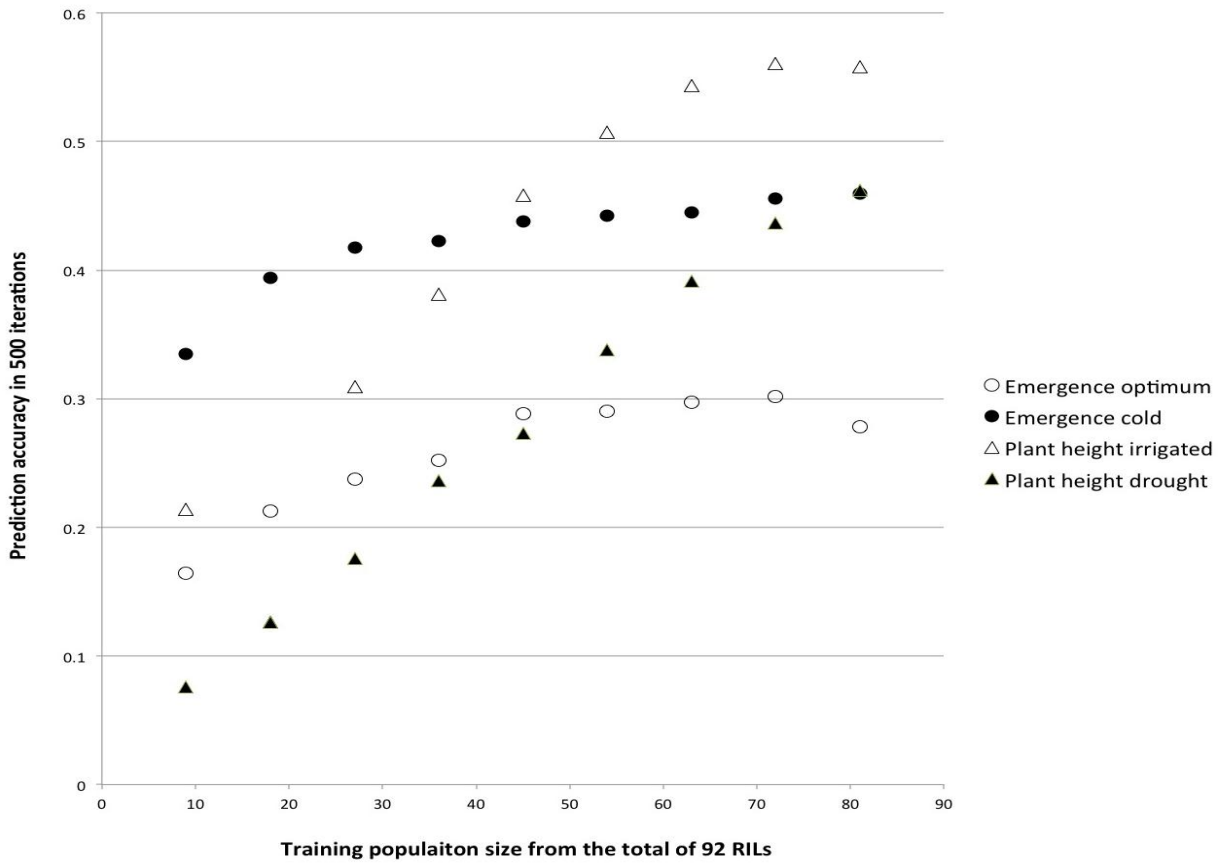


Figure 4.3 Results from genomic prediction and vaidation comparison by taking subgroups of the SM population at 9 different levels of training population size from 9 to 81 of the total 92 RILs used. Mean cross-validation accuracy of the the different trainig population sizes used are shown on the y axis.

Despite the high heritability of emergence both under cold (h^2=0.92) and optimum conditions (h^2=0.91) (Bekele et al. 2013a), the prediction accuracy of the cold emergence was much higher compared to the prediction accuracy for emergence rate under optimum conditions. This could be due to a contrasting genetic architecture of the trait under the different environmental conditions, as shown by the different QTL results in these two conditions (Appendix III, Table 2). Two main effect QTL and eight epistatic QTL for cold emergence explained more than 70% of the phenotypic variance within the 92 lines. On the other hand, the QTL model for emergence under optimal conditions, where only one significant main effect QTL and no epistatic QTL were detected, explained only 12.5% of the total variance. For breeders this means that it should be possible to use genome-wide SNP screening, rather than time-consuming and costly phenotypic screens, for selection of genotypes with a genetic disposition for chilling emergence. Traits like cold emergence are very difficult to phenotype under field conditions due to unpredictable temperature variation from season to season.

A recent comparison between index selection on secondary traits and GS for yield under drought demonstrated a significant prediction advantage of GS over the phenotypic index selection (Ziyomo and Bernardo 2013). In addition to the improvement in genetic gain per breeding cycle, this can potentially further increase breeding progress by allowing an increased number of selection cycles per year. The efficiency of GS in plant breeding has already been tested and optimized for numerous self-pollinated and cross-pollinated crops for various yield, quality and adaptation traits (Heffner et al. 2010; Heffner et al. 2011; Hofheinz et al. 2012). Nevertheless, Jonas and Koning (2013) pointed out the necessity for multiple-cycle cross-validation studies in plants. Furthermore, before GS becomes a routine application in plant breeding it needs to be optimized to cope with genotype by environment variation, which is less of a problem in animal breeding. Heslot et al. (2013) showed that integration of environmental covariances in GS models increased prediction accuracy. Alternative statistical models like Bayes and LASSO may also improve prediction accuracies, however recent examples using wheat and Arabidopsis recommended rr-BLUP as the most suitable model for plant breeding studies (Wimmer et al. 2013).

4.6 Systems biology in plant breeding

Plant breeding has improved global food demand by increasing yields of most major crops by around 1% per annum (Ahlemeyer and Friedt 2012). Genomics-assisted breeding has contributed a small share during the last 20 years (Blum 2013). To meet the demand for even greater increases in yield, caused by the growing world population, in the face of new challenges like climate change and reduced agricultural input, requires innovative new methods to increase selection gains. Systems biological approaches can improve our understanding of complex traits like climatic adaptability and develop multidimensional statistical models which can be deployed to more efficiently develop stable and high yielding varieties.

Systems biology attempts to integrate data from gene to phenotype by dissection of the phenotype at different levels of biological organisation. Collecting comprehensive phenomics data covering all levels of the plant from cell to whole plant does the efficient dissection of a trait. Systems biology fits molecular genetic, biochemical, physiological and imaging technologies together in complex mathematical models and visualization algorithms (Dhont et al. 2013). It requires the incorporation of physiological linkages, trait interactions, and internal plant regulation that can show the patterns of crop growth and development, reflecting genetic control and biological robustness. Although the degree at which such data can be integrated in breeding programs remains to be seen, systems biology analyses are likely to provide breeders with powerful molecular tools to increase their selection gain by improving the accuracy of phenotype prediction and projecting the effect of selection on complex traits interacting with the environment (Cooper et al. 2009). Applications have recently been published for wheat, pea and rice: For example, Wang et al. (2013) recently conducted a systems analysis using microarray data coupled with QTL studies to understand the mechanisms of salt tolerance in rice roots. This led to the identification of interesting candidate genes within QTL support intervals. The present work attempted to dissect early stage chilling stress response of sorghum using mainly morpho-physiological data. Integration of additional layers of metabolome and transcriptome data from carefully designed physiological experiments is the natural continuation of this quest to understand the overriding biological system controlling adaptation of sorghum to early stage chilling and other abiotic stress conditions.

4.7 Conclusions

The plant materials, genetic data and genomics tools presented in this thesis represent valuable resources for breeding for adaptation of sorghum to central European conditions. In particular, they provide a basis for marker-assisted selection (MAS) and genomic selection (GS) for important adaptation and energy traits. Pragmatic and effective utilization of MAS and GS is a prerequisite for the design of efficient breeding programs. For example, MAS for maturity or plant height can be combined with GS for cold emergence. However as shown for cold emergence, plant response and adaptation to abiotic stress are complex processes controlled by multiple regulatory mechanisms. Systems-based models implementing metabolomic and transcriptomic data, and their interaction with variable environments, can give insights into the molecular basis of abiotic stress responses and adaptation of the ecologically versatile crop plant sorghum.

5 Summary

Sorghum (*Sorghum bicolor* (L.) Moench) is the world's 5th important cereal crop, today being grown worldwide in various forms as a food, feed and bioenergy crop. Sorghum has a very wide agro-climatic adaptation due to its high phenotypic diversity, which enabled its expansion from West Africa, via the dryland areas of the Sahel into temperate regions of China and Central America. Its unusually high tolerance of abiotic stress makes it particularly attractive for agroecosystems with low or erratic rainfall. Sorghum has a small, diploid genome (2n=20) which has been completely sequenced, making it a highly suitable model plant for closely related bioenergy crops like sugarcane and Miscanthus.

The major goal of plant production is finding the right crop that can meet our demand for food, feed and fuel without damaging the environment. Maize, the world's most successful multi-purpose crop, is the number one summer crop in many European countries including Germany. The high increase of the maize production area is a leading current topic dominating environmental and agricultural-political discussions in Germany. Sorghum production can readily substitute maize and potentially mitigate some of the problems associated with bio-energy maize production. Sorghum is one of the hardiest plants, with an efficient C4 photosynthetic system and resistance to the maize western rootworm, a devastating pest. It has lower nitrogen and phosphorous demand and can achieve biomass yields that are competitive to maize. Especially sweet sorghum types have the potential to be used for bio-ethanol in addition to biogas production, although storability of the sugar needs to be improved.

On the other hand, however, sorghum is sensitive to early stage chilling stress and is a short-day plant; these two adaptation constraints currently hinder its expansion into temperate agro-ecosystems. Fortunately, there is ample variability in sorghum for many traits including cold tolerance, especially in lines from tropical highland areas. The development of effective breeding strategies for adaptation requires a good understanding of the genetic architecture of the crucial adaptative traits. This study dissected early stage seedling development of sorghum to reveal the complex genetics underlying the slow or retarded growth of sorghum seedlings under chilling stress conditions (Bekele et al. 2013a). Controlled experiments and field trials on a recombinant inbred line (RIL) population from the cross between the sweet sorghum parent SS79 and the grain sorghum parent M71 showed contrasting segregation for pre and post emergence chilling stress. In general, chilling stress reduces emergence, root and

seedling establishment. When the stress is sustained for a long time it causes reduced survival and ultimately death in genotypes that have insufficient chilling tolerance. Interrelationships/correlations among a large number of complex traits were confirmed by the co-location of QTL for multiple traits including emergence, root development and survival under prolonged chilling stress. Highly interesting QTL colocalization hubs were identified on sorghum chromosomes Sb06 and Sb01. Genome re-sequencing of the parental lines and comparison to other selected grain and sweet sorghum genotypes identified more than 1 million single nucleotide polymorphisms (SNPs). The patterns of polymorphisms identified interesting regions which might be of interest to understand the genetic changes that gave rise to sweet and grain morpho-types of sorghum. Next-generation sequencing-based bulk-segregant analysis, on 60 lines showing the highest and the lowest survival under chilling conditions, identified around 7000 SNPs that were unique to either the chilling-susceptible or chilling-tolerant phenotype group. A 3000-SNP Illumina genotyping array was developed for genetic analyses of chilling stress responses using a stringent selection of the genome-wide and trait-linked SNPs (Bekele et al. 2013b).

The SNP array was used to screen a total of 564 sorghum lines, consisting of segregating mapping populations and a diversity panel. The genotype data was used for genome wide association mapping of emergence and for biparental QTL mapping of multiple traits including field biomass, chlorophyll content and brix (sugar content). This efficiently mapped major QTL to known major genes, and in other cases enabled identification of interesting candidate genes for previously unknown QTL.

As a proof-of-concept, the SNP array data was used to test genome-wide prediction (genomic selection) for selected traits using the widely-used ridge-regression best linear unbiased prediction (rr-BLUP) model. Using rr-BLUP with even very small training and validation populations it was possible to detect emergence and plant height under stress and optimum conditions at good cross-validation accuracy of 0.30-0.55. This opens the possibility to use genomic prediction for recurrent selection in breeding programs for difficult traits like chilling emergence. Prediction accuracies will be improved by the use of alternative selection models and the design of breeding strategies account for the specific genetic architecture of each trait. The potential of systems biology in sorghum adaptation research for identification of key regulatory genes is discussed in the context of its potential impact on plant breeding. In the long term the integration of additional levels of data (transcriptome, metabolome) can potentially further improve the selection accuracy of genomic selection.

6 Zusammenfassung

Sorghum (*Sorghum bicolor* (L.) Moench) ist die fünft wichtigste Getreidepflanze der Welt, die als Nahrungs-, Futter- und Bioenergiepflanze angebaut wird. Aufgrund ihrer großen phänotypischen Diversität verfügt Sorghum über eine breite agroklimatische Anpassung. Diese erlaubt einen Anbau von West-Afrika über die Trockengebiete der Sahel-Zone bis in die gemäßigten Klimazonen Chinas und Zentral-Amerikas. Die ungewöhnlich hohe Toleranz gegenüber abiotischem Stress macht Sorghum auch interessant für Agrarsysteme mit geringem bzw. unregelmäßigem Niederschlag. Aufgrund des kleinen, diploiden Genoms (2n=20), das vollständig sequenziert wurde, kann Sorghum als Modellpflanze für eng verwandte Bioenergiepflanzen wie Zuckerrohr und Miscanthus dienen.

Das wichtigste Ziel der Pflanzenproduktion ist die Identifizierung einer Kulturart, die sowohl für die menschliche Ernährung, als auch als Futter- und energieliefernde Pflanze verwendet werden kann, ohne die Umwelt zu schädigen. Mais ist die weltweit erfolgreichste Kulturart und wird aufgrund seiner vielseitigen Nutzung auch vorrangig in vielen europäischen Ländern - einschließlich Deutschlands - angebaut. Der starke Anstieg der Maisanbaufläche ist jedoch ein Hauptthema der umwelt- und agrarpolitischen Diskussionen in Deutschland. Sorghum stellt hingegen eine geeignete Alternative zu Mais dar und könnte damit möglicherweise einige der Probleme verringern, die mit der Produktion von Energie-Mais verbunden sind. Sorghum verfügt über ein wassersparendes C4-Photosynthese-System und gleichzeitig über eine Resistenz gegen den Maiswurzelbohrer, einem verheerenden Schädling. Sorghum kann auch bei einem geringeren Stickstoff- und Phosphorbedarf Biomasseerträge erreichen, die konkurrenzfähig zu Mais sind. Besonders Zuckerhirse-Typen haben das Potential, sowohl für die Bioethanol- als auch für die Biogasproduktion verwendet zu werden.

Andererseits ist Sorghum jedoch empfindlich gegenüber Kühlestress in frühen Pflanzenentwicklungsphasen und ist eine Kurztagspflanze. Diese Einschränkungen verhindern momentan eine weitere Ausdehnung des Anbaus in gemäßigtere Agrar-Ökosysteme. Glücklicherweise besitzt Sorghum für viele Merkmale eine große genetische Variabilität. Interessante Variation z.B. für Kühletoleranz findet man in Sorghum-Linien aus den tropischen Hochländen. Die Entwicklung effektiver Züchtungsstrategien für die Verbesserung dieses Merkmals erfordert jedoch ein gutes Verständnis der ihnen zugrundeliegenden genetischen Struktur.

Im Rahmen dieser Studie werden die frühen Entwicklungsphasen des Keimlings analysiert, um die komplexe Genetik aufzudecken, die das langsame bzw. verzögerte Wachstum von Sorghum-Keimlingen unter Kühlestress-Bedingungen bewirkt. Experimente unter kontrollierten Wachstumsbedingungen sowie Feldversuche mit rekombinanten Inzuchtlinien (RIL)-Populationen aus Kreuzungen zwischen dem Zuckerhirse-Elter SS79 und dem Körnerhirse-Elter M71 zeigten gegensätzliche Reaktionen auf Kältestress vor oder nach Keimlingsaufgang. Generell reduzierte Kühlestress den Keimlingsaufgang, die Wurzel- und die Keimlingsentwicklung. Wenn der Stress über einen längeren Zeitraum anhielt, bewirkte er ein verringertes Überleben und schließlich das Absterben der Genotypen mit einer unzureichenden Kühletoleranz.

Die Zusammenhänge bzw. Korrelationen zwischen einer großen Anzahl an komplexen Merkmalen konnten mittels der Kolokalisierung von QTL für mehrere Merkmale wie Aufgang, Wurzelentwicklung und Überleben unter anhaltendem Kühlestress aufgedeckt werden. Besonders interessante Bereiche mit kolokalisierenden QTL wurden auf den Chromosomen Sb06 und Sb01 identifiziert. Die Resequenzierung des Genoms der elterlichen Linien und der Vergleich mit anderen ausgewählten Körner- und Zuckerhirse-Genotypen ergab mehr als eine Million Einzelbasenpolymorphismen (*Single Nucleotide Polymorphisms*, SNPs). Die SNP-Verteilungsmuster konnten auch dabei helfen, die genetischen Veränderungen zu verstehen, die zu der Entwicklung von Zucker- und Körner-Morphotypen in Sorghum führten. Eine auf *Next-Generation*-Sequenzierung basierte *Bulk-Segregant*-Analyse mit 60 Linien, welche die höchste bzw. die niedrigste Überlebensrate unter Kältestress zeigten, identifizierte ca. 7000 SNP-Marker, die entweder einzigartig in der Gruppe der kälteempfindlichen oder in der Gruppe der kühletoleranten Phänotypengruppe gefunden wurden. Eine stringente Auswahl von 3000 genomweiten bzw. merkmalsassoziierten SNPs wurde daraufhin zur Entwicklung eines *Illumina-Genotyping-Array* für genetische Analysen der Kältestressantworten eingesetzt.

Mit dem 3000-SNP-*Array* konnten insgesamt 564 Sorghum-Linien aus unterschiedlichen Kartierungspopulationen sowie einem Diversitätsset untersucht werden. Die Genotypendaten wurden für eine genomweite Assoziationskartierung des Keimlingsaufgangs sowie für die QTL-Kartierung der Merkmale Feldbiomasse, Chlorophyllgehalt und *BRIX* (Zuckergehalt) der beiden Eltern verwendet. Damit konnten einerseits erfolgreich Major-QTL kartiert werden, die in Zusammenhang mit bereits bekannten Genen stehen. In anderen Fällen wurden auch interessante, neue Kandidatengene für bisher unbekannte QTL identifiziert. Solche Gene

bzw. eng gekoppelte SNP-Marker stellen einen wichtigen Durchbruch für die Züchtung auf Kühletoleranz dar.

Als *Proof-of-Concept* wurden die Daten des SNP-*Array* benutzt, um die genomische Selektion für bestimmte Merkmale mittels des häufig benutzten *Ridge-Regression Best Linear Unbiased Prediction* (rr-BLUP)-Modells zu überprüfen. Mit Hilfe von rr-BLUP war es sogar in sehr kleinen Übungs- und Validierungs-Populationen möglich, Keimlingsaufgang und Pflanzenhöhe unter optimalen sowie unter Stressbedingungen mit überraschend guter Kreuzvalidierungsgenauigkeit von 0,30-0,55 zu detektieren. Dies ermöglicht die genomische Vorhersage für die rekurrente Selektion in Zuchtprogrammen für schwierige Merkmale wie Kühletoleranz. Die Vorhersagegenauigkeit kann voraussichtlich durch die Benutzung alternativer Selektionsmodelle und Züchtungsstrategien, welche die spezifische genetische Struktur jedes Merkmals berücksichtigen, noch weiter gesteigert werden. Langfristig könnten die züchterische Selektion bzw. das genetische Verständnis der Kühletoleranz in Sorghum durch die Integration von weiteren Datenebenen (z.B. Transkriptom, Metabolom) potentiell noch weiter verbessert werden.

7 References

Adugna A, Snow AA, Sweeney PM, Bekele E and Mutegi E (2013) Population genetic structure of in situ wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers. *Genetic Resources and Crop Evolution* 60, 1313-1328.

Ahleyer J and Friedt W (2012) Winterweizenerträge in Deutschland stabil auf hohem Niveau – Welchen Einfluss hat der Züchtungsfortschritt? *Getreidemagazin* 17, 38-41.

Aroca R, Tognoni F, Irigoyen JJ, Sánchez-Diaz M and Pardossi A (2001) Different root low temperature response of two maize genotypes differing in chilling sensitivity. *Plant Physiology & Biochemistry* 39, 1067–1073.

Bekele WA, Fiedler K, Shiringani A, Schnaubelt D, Windpassinger S, Uptmoor R, Friedt W and Snowdon R (2013a) Unravelling the genetic complexity of sorghum seedling development under low-temperature conditions. *Plant, Cell & Environment* doi: 10.1111/pce.12189.

Bekele WA, Wieckhorst S, Friedt W and Snowdon R (2013b) High-throughput genomics in sorghum: from whole-genome resequencing to a SNP screening array development. *Plant Biotechnology Journal* 11, 1112-1125.

Bhosale SU, Rymen B, Beemster GT, Melchinger AE and Reif JC (2007) Chilling tolerance of Central European maize lines and their factorial crosses. *Annals of Botany* 100,1315-1321.

Blum A (2011) Drought resistance – is it really a complex trait? *Functional Plant Biology* 38, 753-757.

Blum A (2013) Heterosis, stress, and the environment: a possible road map towards the general improvement of crop yield. *Journal of Experimental Botany* doi: 10.1093/jxb/ert289

Borrell AK and Hammer GL (2000) Nitrogen dynamics and the physiological basis of stay-green in sorghum. *Crop Science* 40, 1295-1307.

Borrell AK, Hammer GL and Douglas ACL (2000) Does maintaining green leaf area in sorghum improve yield under drought? I. Leaf growth and senescence. *Crop Science* 40, 1026-1037.

Boyer JS (1982) Plant productivity and environment. *Science* 218, 443–448.

Brar GS, Steiner JL, Unger PW and Prihar SS (1992) Modelling sorghum seedling establishment from soil wetness and temperature of drying seed zones. *Agronomy Journal* 84, 905–910.

Brown PJ and Paterson HA (2013) The gene pool of *Sorghum bicolor* and its improvement. In Genomics of the Saccharinae (pp 333-346), Plant Genetics and Genomics: Crops and Models 11 Paterson AH (ed.) New York, Springer.

Brown PJ, Myles S and Kresovich S (2011) Genetic support for phenotype-based racial classification in sorghum. *Crop Science* 51, 224-230.

Buckler ES and Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* 5, 107–111.

Burow G, Burke JJ, Xin Z and Franks CD (2011) Genetic dissection of early-season cold tolerance in sorghum (*Sorghum bicolor* (L.) Moench). *Molecular Breeding* 28, 391–402.

Cai H, Dunn JB, Wang Z, Han J and Wang MQ (2013) Life-cycle energy use and greenhouse gas emissions of production of bioethanol from sorghum in the United States. *Biotechnology for Biofuels* 6,141 doi:10.1186/1754-6834-6-141.

Chinnusamy V, Zhu J and Zhu JK (2007) Cold stress regulation of gene expression in plants. *Trends in Plant Science* 12, 444–451.

Claeys H and Inzé D (2013) The agony of choice: how plants balance growth and survival under water-limiting conditions. *Plant Physiology* 162, 1768-1779.

Collard BCY, Jahufer MZZ, Brouwer JB and Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196.

Cooper M, van Eeuwijk FA, Hammer GL and Podlich DW (2009) Modelling QTL for complex traits: detection and context for plant breeding. *Current Opinion in Plant Biology* 12, 231-240.

Cramer GR, Urano K, Delrot S, Pezzotti M and Shinozaki K (2011) Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biology* 11, 163.

Craufurd PQ, Mahalakshmi V, Bidinger FR, Mukuru SZ, Chantereau J, Omanga PA, Qi A, Roberts EH, Ellis RH, Summerfield RJ and Hammer GL (1999) Adaptation of sorghum: characterisation of genotypic flowering responses to temperature and photoperiod. *Theoretical and Applied Genetics* 99, 900–911.

Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D and Mathews K (2013) Genomic prediction in CIMMYT maize and wheat breeding programs. *Heredity* doi: 10.1038/hdy.2013.16.

Cruz de Carvalho MH (2008) Drought stress and reactive oxygen species: Production, scavenging and signaling. *Plant Signaling & Behavior* 3,156-165.

Dahlberg J, Berenji J, Sikora V and Latković D (2011) Assessing sorghum [*Sorghum bicolor* (L.) Moench] germplasm for new traits: food, fuels & unique uses. *Maydica* 56, 85-92.

Dahlberg JA and Wasylikowa K (1996) Image and statistical analyses of early sorghum remains (8000 B.P.) from the Nabata Playa archaeological site in the Western Desert, southern Egypt. *Vegetation History and Arachaeobotany* 5, 293-299.

de Wet JMJ (1978) Systematics and evolution of sorghum sect. Sorghum (gramineae). *American Journal of Botany* 65, 477–484.

Des Marais DL and Juenger TE (2010) Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. *Annals of the New York Academy of Sciences* 1206, 56-79.

Dhont S, Wuyts N and Inze D (2013) Cell to whole-plant phenotyping: the best is yet to come. *Trends in Plant Science* 16, 428-439.

Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L and Lee S (2007) Domestication to crop improvement: Genetic resources for *Sorghum* and *Saccharum* (Andropogoneae). *Annals of Botany* 100, 975-989.

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250-255.

Endelman JB and Jannink JL (2012) Shrinkage estimation of the realized relationship matrix. *Genes, Genomes, Genetics* 2, 1405-1413.

Edwards EJ and Smith SA (2010) Phylogenetic analyses reveal the shady history of C4 grasses. *Proceedings of the National Academy of Sciences of the United States of America* 107, 2532–2537.

Ejeta G (2007) Breeding for striga resistance in sorghum: Exploitation of an intricate host–parasite biology. *Crop Science* 47, 216-217.

Falconer DS (1989) Quantitative Genetics Longman, Scientific & Technical (Burnt Mill, Harlow, Essex, England and New York).

FAO's Director-General (2009) FAO's Director-General on How to Feed the World in 2050. *Population and Development Review* 35, 837-839.

Fiedler K, Bekele WA, Friedt W, Snowdon R, Stuetzel H, Zacharias A and Uptmoor R (2012) Genetic dissection of temperature-dependent emergence processes in sorghum using a cumulative emergence model and stability parameters. *Theoretical and Applied Genetics* 125, 1647-1661.

Fitzgerald TL, Shapter FM, McDonald S, Waters DLE, Chivers IH, Drenth A, Nevo E and Henry RJ (2011) Genome diversity in wild grasses under environmental stress. *Proceedings of the National Academy of Sciences of the United States of America* 108, 21140-21145.

Franks C, Burow G and Burke J (2006) A comparison of U.S. and Chinese sorghum germplasm for early season cold tolerance. *Crop Science* 46, 1371-1376.

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner E-M, Hansen M, Joets J, Le Paslier M-C, McMullen MD, Montalent P, Rose M, Schön C-C, Sun Q, Walter H, Martin OC and Falque M (2011) A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PLoS ONE* 6, e28334.

Gaspar T, Franck T, Bisbis B, Kevers C, Jouve L, Hausman JF and Dommes J (2002) Concepts in plant stress physiology. Application to plant tissue cultures. *Plant Growth Regulation* 37, 263–285.

Gouy M, Rousselle Y, Bastianelli D, Lecomte P, Bonnal L, Roques D, Efile JC, Rocher S, Daugrois J, Toubi L, Nabeneza S, Hervouet C, Telismart H, Denis M, Thong-Chane A, Glaszmann JC, Hoarau JY, Nibouche S and Costet L (2013) Experimental assessment of the accuracy of genomic selection in sugarcane. *Theoretical and Applied Genetics* 126, 2575-2586.

Hammer GL, Dong Z, McLean G, Doherty A, Messina C, Schussler J, Zinselmeier C, Paszkiewicz A and Cooper M (2009) Can changes in canopy and/or root system

architecture explain historical maize yield trends in the U.S. corn belt? *Crop Science* 49, 299-312.

Harlan JR (1971) Agricultural origins: Centres and non centres. *Science* 174, 468-474.

Harlan JR and de Wet JMJ (1972) A simplified classification of cultivated sorghum. *Crop Science* 12, 172–176.

Harris K, Subudhi PK, Borrell A, Jordan D, Rosenow D, Nguyen H, Klein R and Mullet J (2007) Sorghum stay-green QTL individually reduce post-flowering drought induced senescence. *Journal of Experimental Botany* 58, 327–338.

Hayes BJ and Goddard ME (2010) Genome-wide association and genomic selection in animal breeding. *Genome* 53, 876–883.

Heffner EL, Jannink JL, Iwata H, Souza E and Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop Science* 51, 2597–2606.

Heffner EL, Lorenz AJ, Jannink JL and Sorrells ME (2010) Plant breeding with genomic selection: Gain per unit time and cost. *Crop Science* 50, 1681-1690.

Heffner EL, Sorrells ME and Jannink JL (2009) Genomic selection for crop improvement. *Crop Science* 49, 1-12.

Heslot N, Akdemir D, Sorrells ME and Jannink JL (2013) Integrating environmental convariates and crop modelling into the genomic selection framework to predict genotype by environment interaction. *Theoretical and Applied Genetics* Epub ahead of print.

Heslot N, Yang HP, Sorrells ME and Jannink JL (2012) Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science* 52, 146-160.

Heyndrickx KS and Vandepoele K (2012) Systematic identification of functional plant modules through the integration of complementary data sources. *Plant Physiology* 159, 884-901.

Hibberd JM, Sheehy JE and Langdale JA (2008) Using C4 photosynthesis to increase the yield of rice-rationale and feasibility. *Current Opinion in Plant Biology* 11, 228-231.

Hofheinz N, Borchardt D, Weissleder K and Frisch M (2012) Genome-based prediction of test cross performance in two subsequent breeding cycles. *Theoretical and Applied Genetics* 125, 1639–1645.

Holland, JB (2007) Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology* 10, 156-161.

Huang X, Feng Q, Qian Q, Zhao Q, Wang L, Wang A, Guan J, Fan D, Weng Q, Huang T, Dong G, Sang T and Han B (2009) High-throughput genotyping by whole-genome resequencing. *Genome Research* 19, 1068–1076.

Hund A, Frachebound Y, Soldati A and Frascaroli E (2004) QTL controlling root and shoot traits of maize seedlings under cold stress. *Theoretical and Applied Genetics* 109, 618–629.

Jaggard KW, Qi A and Ober ES (2010) Possible changes to arable crop yield by 2050. *Philosophical Transactions of the Royal Society Biological Sciences* 365, 2835-2851.

Jannink JL, Lorenz AJ and Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Briefings in Functional Genomics* 9, 166–177.

Jessup RW (2013) Perennialism and weediness in the Saccharinae. In Genomics of the Saccharinae (pp 503-518), Plant Genetics and Genomics: Crops and Models Paterson AH (ed.) New York, Springer.

Jiang SY, Ma Z, Vanihta J and Ramachandran S (2013) Genetic variation and expression diversity between grain and sweet sorghum lines. *BMC Genomics* 14, 18.

Jonas E and de Koning DJ (2013) Does genomic selection have a future in plant breeding? *Trends in Biotechnology* 31, 497-504.

Jordan DR, Mace ES, Cruickshank AW, Hunt CH and Henzell RG (2011) Exploring and exploiting genetic variation from unadapted sorghum germplasm in a breeding program. *Crop Science* 51, 1444–1457.

Juenger TE (2013) Natural variation and genetic constraints on drought tolerance. *Current Opinion in Plant Biology* 16, 274-281.

Kebede H, Subudhi PK, Rosenow DT and Nguyen HT (2001) Quantitative trait loci influencing drought tolerance in grain sorghum (*Sorghum bicolor* (L.) Moench). *Theoretical and Applied Genetics* 103, 266–276.

Kellog EA (2013) The gene pool of *Sorghum bicolor* and its improvement. In Genomics of the Saccharinae (pp 3-22), Plant Genetics and Genomics: Crops and Models Paterson AH (ed.) New York, Springer.

Kimber CT, Dahlberg JA and Kresovich S (2013) The gene pool of *Sorghum bicolor* and its improvement. In Genomics of the Saccharinae (pp 23-41), Plant Genetics and Genomics: Crops and Models Paterson AH (ed.) New York, Springer.

Kloosterman B, Abelenda JA, Gomez Mdel M, Oortwijn M, de Boer JM, Kowitwanich K, Horvath BM, van Eck HJ, Smaczniak C, Prat S, Visser RG and Bachem CW (2013) Naturally occurring allele diversity allows potato cultivation in northern latitudes. *Nature* 14, 246-250.

Knoll J and Ejeta G (2008) Marker-assisted selection for early-season cold tolerance in sorghum: QTL validation across populations and environments. *Theoretical and Applied Genetics* 116, 541–553.

Knoll J, Gunaratna N and Ejeta G (2008) QTL analysis of early-season cold tolerance in sorghum. *Theoretical and Applied Genetics* 116, 577–587.

Knox AK, Dhillon T, Cheng H, Tondelli A, Pecchioni N and Stockinger EJ (2010) CBF gene copy number variation at Frost Resistance-2 is associated with levels of freezing tolerance in temperate-climate cereals. *Theoretical and Applied Genetics* 121, 21-35.

Krasensky J and Jonak C (2012) Drought, salt, and temperature stress-induced metabolic rearrangements and regulatory networks. *Journal of Experimental Botany* 63, 1593-1608.

Kresovich S, Barbazuk B, Bedell JA, Borrell A, Buell CR, Burke J, Clifton S, Cordonnier-Pratt MM, Cox S, Dahlberg J, Erpelding J, Fulton TM, Fulton B, Fulton L, Gingle AR, Hash CT, Huang Y, Jordan D, Klein PE, Klein RR., Magalhaes J, McCombie R, Moore P, Mullet JE, Ozias-Akins P, Paterson AH, Porter K, Pratt L, Roe B, Rooney W, Schnable PS, Stelly DM, Tuinstra M, Ware D and Warek U (2005) Toward sequencing the sorghum genome. A US National Science Foundation-sponsored Workshop Report. *Plant Physiology* 138, 1898-1902.

Krothapalli K, Buescher EM, Li X, Brown E, Chapple C, Dilkes BP and Tuinstra MR (2013) Forward genetics by genome sequencing reveals that rapid cyanide release deters insect herbivory of *Sorghum bicolor*. *Genetics* 195, 309-318.

KWS Saat AG (2013) KWS Energiesorghum für 2013. Energy sorghum production manual for year 2013.

Lander ES and Schork NJ (1994) Genetic dissection of complex traits. *Science* 265, 2037-2048.

Lawrence CJ and Walbot V (2007) Translational genomics for bioenergy production from fuelstock grasses: Maize as the model species. *The Plant Cell* 19, 2091-2094.

Leipner J (2009) Chilling stress in maize: From physiology to genetics and molecular mechanisms. Habilitation Thesis. Eidgenössische Technische Hochschule Zürich, 124p.

Leipner J, Fracheboud Y and Stamp P (1997) Acclimation by suboptimal growth temperature diminishes photooxidative damage in maize leaves. *Plant, Cell & Environment* 20, 366–372.

Li L, Liu X, Xie K, Wang Y, Liu F, Lin Q, Wang W, Yang C, Lu B, Liu S, Chen L, Jiang L and Wan J (2013) *qLTG-9*, a stable quantitative trait locus for low-temperature germination in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics* 126, 2313-2322.

Li M, Yuyama N, Luo L, Hirata M and Cai H (2009) *In silico* mapping of 1758 new SSR markers developed from public genomic sequences for sorghum. *Molecular Breeding* 24, 41–47.

Lin Z, Li X, Shannon LM, Yeh CT, Wang ML, Bai G, Peng Z, Li J, Trick HN, Clemente TE, Doebley J, Schnable PS, Tuinstra MR, Tesso TT, White F and Yu J (2012) Parallel domestication of the *Shattering1* genes in cereals. *Nature Genetics* 44, 720-724.

Lorenz, AJ, Smith KP and Jannink JL (2012) Potential and optimization of genomic selection for Fusarium head blight resistance in six-row barley. *Crop Science* 52, 1609–1621.

Lu Y, Zhang S, Shah T, Xie C, Hao Z and Li X (2010) Joint linkage–linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proceedings of the National Academy of Sciences of the United States of America* 107, 19585–19590.

Ma XF, Jensen E, Alexandrov N, Troukhan M, Zhang LP, Thomas-Jones S, Farrar K, Clinton-Brown J, Donnison I, Swaller T and Flavell R (2012) High resolution genetic mapping by genome sequencing reveals genome duplication and tetraploid genetic structure of the diploid *Miscanthus sinensis*. *PLoS ONE* 7, e33821.

Mace ES and Jordan DR (2011) Integrating sorghum whole genome sequence information with a compendium of sorghum QTL studies reveals uneven distribution of QTL and of

gene-rich regions with significant implications for crop improvement. *Theoretical and Applied Genetics* 123, 169–191.

Mace ES, Singh V, Van Oosterom EJ, Hammer GL, Hunt CH and Jordan DR (2012) QTL for nodal root angle in sorghum (*Sorghum bicolor* L. Moench) co-locate with QTL for traits associated with drought adaptation. *Theoretical and Applied Genetics* 124, 97–109.

Mace ES, Tai S, Gilding EK, Li Y, Prentis PJ, Bian L, Campbell BC, Hu W, Innes DJ, Han X, Cruickshank A, Dai C, Frère C, Zhang H, Hunt CH, Wang X, Shatte T, Wang M, Su Z, Li J, Lin X, Godwin ID, Jordan DR and Wang J (2013) Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nature Communications* 4, 2320. doi: 10.1038/ncomms3320.

Maron LG, Guimaraes CT, Kirst M, Albert PS, Birchler JA, Bradbury PJ, Buckler ES, Coluccio AE, Danilova TV, Kudrna D, Magalhaes JV, Peñeros MA, Schatz MC, Wing RA and Kochian LV (2013) Aluminium tolerance in maize is associated with higher MATE1 gene copy number. *Proceedings of the National Academy of Sciences of the United States of America* 110, 5241–5246.

Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, Mizuno H, Yamamoto K, Antonio BA, Baba T, et al. (2005) The map-based sequence of the rice genome. *Nature* 436, 793–800.

McClintock B (1951) Mutable loci in maize. *Carnegie Institution of Washington Yearbook* 50, 174-181.

McCouch S, Baute GJ, Bradeen J, Bramel P, Bretting PK, Buckler E, Burke JM, Charest D, Cloutier S, Cole G, Dempewolf H, Dingkuhn M, Feuillet C, Gepts P, Grattapaglia D, Guarino L, Jackson S, Knapp S, Langridge P, Lawton-Rauh A, Lijua Q, Lusty C, Michael T, Myles S, Naito K, Nelson RL, Pontarollo R, Richards CM, Rieseberg L, Ross-Ibarra J, Rounsley S, Hamilton RS, Schurr U, Stein N, Tomooka N, van der Knaap E, van Tassel D, Toll J, Valls J, Varshney RK, Ward J, Waugh R, Wenzl P and Zamir D (2013) Agriculture: Feeding the future. *Nature* 499, 23-24.

Meuwissen THE, Hayes BJ and Goddard ME (2001) Prediction of total genetic value using genome wide dense marker maps. *Genetics* 157, 1819-1829.

Michael TP and Jackson S (2013) The first 50 plant genomes. *The Plant Genome* 6, 1-7.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE, Harriman J, Glaubitz JC, Buckler ES and Kresovich S (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, 453–458.

Multani DS, Briggs SP, Chamberlin MA, Blakeslee JJ, Murphy AS and Johal GS (2003) Loss of an MDR transporter in compact stalks of maize *br2* and sorghum *dw3* mutants. *Science* 302, 81-84.

Munkvold JD, Laudencia-Chingcuanco D and Sorrells ME (2013) Systems genetics of environmental response in the mature wheat embryo. *Genetics* 194, 265-277.

Muñoz-Amatriain M, Eichten SR, Wicker T, Richmond TA, Mascher M, Steuernagel B, Scholz U, Ariyadasa R, Spannagl M, Nussbaumer T, Mayer KFX, Taudien S, Platzer M,

Jeddeloh JA, Springer NM, Muehlbauer GJ and Stein N (2013) Distribution, functional impact, and origin mechanisms of copy number variation in the barley genome. *Genome Biology* 14, R58.

Muraya MM, Mutegi E, Geiger HH, de Villiers SM, Sagnard F, Kanyenji BM, Kiambi D and Parzies HK (2011a) Wild sorghum from different eco-geographic regions of Kenya display a mixed mating system. *Theoretical and Applied Genetics* 122, 1631-1639.

Muraya MM, de Villiers SM, Parzies HK, Mutegi E, Sagnard F, Kanyenji BM, Kiambi D and Geiger HH (2011b) Genetic structure and diversity of wild sorghum populations (*Sorghum* spp.) from different eco-geographical regions of Kenya. *Theoretical and Applied Genetics* 123, 571-583.

Murphy RL, Klein RR, Morshige DT, Brady JA, Rooney WL, Miller FR, Dugas DV, Klein PE and Mullet JE (2011) Coincident light and clock regulation of *pseudoresponse regulator protein 37 (PRR37)* controls photoperiodic flowering in sorghum. *Proceedings of the National Academy of Sciences of the United States of America* 108, 16469-16474.

Nelson, JC, Wang S, Wu Y, Li X, Antony G, White FF and Yu JM (2011) Single-nucleotide polymorphism discovery by high-throughput sequencing in sorghum. *BMC Genomics* 12, 352.

Nordborg M and Weigel D (2008) Next-generation genetics in plants. *Nature* 456, 720–723.

Nunes C, Schluepmann H, Delatte TL, Wingler A, Silva AB, Fevereio PS, Jansen M, Fiorani F, Wiese-Klinkenberg A and Paul M (2013) Regulation of growth by the trehalose pathway: Relationship to temperature and sucrose. *Plant Signaling & Behavior* 8, e26626.

Paterson AH (2008) Genomics of sorghum. *International Journal of Plant Genomics* 2008, 362–451.

Paterson AH (2013) Preface. In Genomics of the Saccharinae (pp vii-viii), Plant Genetics and Genomics: Crops and Models Paterson AH (ed.) New York, Springer.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J and Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. *Nature* 457, 551–556.

Payne WA, Batata M and Rosenow DT (2003) Sorghum diversity for germination and coleoptile elongation under cool conditions. *International Sorghum and Millets Newsletter* 44, 76–79.

Peacock JM (1982) Response and tolerance of sorghum to temperature stress. Sorghum in the eighties: Proceedings of the International Symposium on Sorghum (pp 143-159), 2–7 November 1981, Patancheru. A.P. India.

Poland JJ, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sánchez-Villeda H, Sorrells M and Jannink JL (2012) Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103-113.

Premachandra GS, Hahn DT, Axtell JD and Joly RJ (1994) Epicuticular wax load and water-use efficiency in bloomless and sparse-bloom mutants of *Sorghum bicolor* L. *Environmental and Experimental Botany* 34, 293-301.

Prochnik S, Marri PR, Desany B, Rabinowicz PD and Kodira C (2012) The cassava genome: Current progress, future directions. *Tropical Plant Biology* 5, 88-94.

Prohens J, Rodriguez-Burruezo A and Nuez F (2007) Introduction and adaptation of new crops. In Principles of plant genetic and breeding (pp 20-22). George Acquaah (ed.) Maiden, MA USA, Blackwell publishing.

Qin F, Shinozaki K and Yamaguchi-Shinozaki K (2011) Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant & Cell Physiology* 52, 1569-1582.

Ray DK, Mueller ND, West PC and Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. *PLoS ONE* 8, e66428.

Riedelsheimer C, Czedik-Eysenberg A, Grieder C, Lisec J, Technow F, Sulpice R, Altmann T, Stitt M, Willmitzer L and Melchinger AE (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. *Nature Genetics* 44, 217–220.

Riedelsheimer C, Endelman JB, Stange M, Sorrells ME, Jannink JL and Melchinger AE (2013) Genomic predictability of interconnected biparental maize populations. *Genetics* 194, 493-503.

Rosenow DT (1977) Breeding for lodging resistance in sorghum. In Proceedings of 32nd Annual Corn and Sorghum Industry Research Conference (pp 171–185) Loden HD and Wilkinson D (eds.). Washington, DC, USA. American Seed Trade Association.

Roy SJ, Tucker EJ and Tester M (2011) Genetic analysis of abiotic stress tolerance in crops. *Current Opinion in Plant Biology* 14, 232–239.

Salih AA, Ali IA, Lux A, Luxova M, Cohen Y, Sugimoto Y and Inanaga S (1999) Rooting, water uptake, and xylem structure adaptation to drought of two sorghum cultivars. *Crop Science* 39, 168-173.

Saze H, Shiraishi A, Miura A and Kakutani T (2008) Control of genic DNA methylation by a jmjC domain-containing protein in *Arabidopsis thaliana*. *Science* 319, 462-465.

Schittenhelm S and Schroetter S (2013) Comparison of drought tolerance of maize, sweet sorghum and sorghum-sudangrass hybrids. *Journal of Agronomy and Crop Science* DOI: 10.1111/jac.12039.

Schneeberger K, Ossowski S, Lanz C, Juul T, Petersen AH, Nielsen KL, Jorgensen JE, Weigel D and Andersen SU (2009) SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* 6, 550-551.

Setimela P, Manthe CS, Mazhani L and Obilana AB (1997) Release of three grain sorghum pure line varities in Botswana. *South African Journal of Plant and Soil* 14, 137-138.

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

Shiringani AL and Friedt W (2011) QTL for fibre-related traits in grain × sweet sorghum as a tool for the enhancement of sorghum as a biomass crop. *Theoretical and Applied Genetics* 123, 999-1011.

Singh V, van Oosterom EJ, Jordan DR, Messina CD, Cooper M and Hammer GL (2010) Morphological and architectural development of root systems in sorghum and maize. *Plant and Soil* 333, 287–299.

Souza GM, Berges H, Bocs S, Casu R, D'Hont A, Ferreira JE, Henry R, Ming R, Potier B, Sluys MA, Vincentz M and Paterson AH (2011) The Sugarcane genome challenge: Strategies for sequencing a highly complex genome. *Tropical Plant Biology* 4, 145.

Stefaniaka TR, Dahlbergb JA, Beanc BW, Dighed N, Wolfrume EJ and Rooney WL (2012) Variation in biomass composition components among forage, biomass, sorghum-sudangrass, and sweet sorghum types. *Crop Science* 52, 1949-1954.

Stephens JC and Holland RF (1954) Cytoplasmic male sterility for sorghum seed production. *Agronomy Journal* 46, 20–23.

Strigens A, Freitag NM, Gilbert X, Grieder C, Riedelsheimer C, Schrag TA, Messmer R and Melchinger AE (2013) Association mapping for chilling tolerance in elite flint and dent maize inbred lines evaluated in growth chamber and field experiments. *Plant, Cell and Environment* 36, 1871-1887.

Takagi H, Abe A, Kentaro Y, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S and Terauchi R (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *The Plant Journal* 74, 174-183.

Tarumoto I (2011) Thermo-sensitivity and photoperiod sensitivity genes controlling heading time and flower bud initiation in sorghum, *Sorghum bicolor* Moench. *Japan Agricultural Research Quarterly* 45, 1-7.

Tarumoto I, Yanase M, Iwahara Y, Kuzumi Y, Morikawa T and Kasuga S (2003) Inheritance of a thermo-sensitivity gene controlling flower initiation in sorghum. *Breeding Science* 53, 353-357.

Teshome A, Fahrig L, Torrance JK, Lambert JD, Arnason TJ and Baum BR (1999) Maintenance of sorghum (*Sorghum bicolor*, Poaceae) landrace diversity by farmers' selection in Ethiopia. *Economic Botany* 53, 79-88.

Thurber CS, Ma JM, Higgins RH and Brown PJ (2013) Retrospective genomic analysis of sorghum adaptation to temperate-zone grain production. *Genome Biology* 14, R68.

Upadhyaya HD, Pundir RPS, Dwivedi SL, Gowda CLL, Reddy VG and Singh S (2009) Developing a mini core collection of sorghum for diversified utilization of germplasm. *Crop Science* 49, 1769-1780.

Varshney RK, Ribaut JM, Buckler ES, Tuberosa R, Rafalski JA and Langridge P (2012) Can genomics boost productivity of orphan crops? *Nature Biotechnology* 30, 1172–1176.

Vermerris W and Saballos A (2013) Genetic enhancement of sorghum for biomass utilization. In Genomics of the Saccharinae (pp 391-425), Plant Genetics and Genomics: Crops and Models Paterson AH (ed.) New York, Springer.

Vigeland MD, Spannagl M, Asp T, Paina C, Rudi H, Rognli OA, Fjellheim S and Sandve SR (2013) Evidence for adaptive evolution of low-temperature stress response genes in a Pooideae grass ancestor. *New Phytologist* 199, 1060-1068.

Wang J, Chen L, Wang Y, Zhang J, Liang Y and Cu D (2013) A computational systems biology study for understanding salt tolerance mechanism in rice. *PLoS ONE* 8, e64929.

Washburn JD, Murray SC, Burson BL, Klein BR and Jessup R (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, 153-162.

Weston LA, Alsaadawi IS and Baerson SR (2013) Sorghum allelopathy-from ecosystem to molecule. *Journal of Chemical Ecology* 39, 142-153.

Wimmer V, Lehermeier C, Albrecht T, Auinger HJ, Wang Y and Schön CC (2013) Genome-wide prediction of traits with different genetic architecture through efficient variable selection. *Genetics* 195, 573-587.

Würschum T, Reif JC, Kraft T, Janssen G and Zhao Y (2013) Genomic selection in sugar beet breeding populations. *BMC Genetics* 14, 85.

Yang Z, Huang D, Tang W, Zheng Y, Liang K, Cutler AJ and Wu W (2013) Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. *PLoS ONE* 8, e68433.

Yu J and Tuinistra MR (2001) Genetic analysis of seedling growth under cold temperature stress in grain sorghum. *Crop Science* 41, 1438-1443.

Yu J, Holland JB, McMullen MD and Buckler ES (2008) Power analysis of an integrated mapping strategy: nested association mapping. *Genetics* 138, 539–551.

Zhao Y, Gowda M, Würschum T, Longin CF, Korzun V, Kollers S, Schachschneider R, Zeng J, Fernando R, Dubcovsky J and Reif JC (2013) Dissecting the genetic architecture of frost tolerance in Central European winter wheat. *Journal of Experimental Botany* 64, 4453-4460.

Zheng LY, Guo XS, He B, Sun LJ, Peng Y, Dong SS, Liu TF, Jiang S, Ramachandran S, Liu CM and Jing HC (2011) Genome-wide patterns of genetic variation in sweet sorghum and grain sorghum (*Sorghum bicolor*). *Genome Biology* 12, R114.

Ziyomo C and Bernardo R (2013) Drought tolerance in maize: Indirect selection through secondary traits versus genomewide selection. *Crop Science* 53, 1269-1275.

Zou G, Zhai G, Feng Q, Yan S, Wang A, Zhao Q, Shao J, Zhang Z, Zou J, Han B and Tao Y (2012) Identification of QTLs for eight agronomically important traits using an ultra-throughput sequencing in sorghum under contrasting photoperiods. *Journal of Experimental Botany* 63, 5451–5462.

8 Appendices

Appendix I: Electronic supplementary materials from Bekele et al. (2013a)

Wubishet A. Bekele[1], Karin Fiedler[2], Amukelani Shiringani[1], Daniel Schnaubelt[1,3], Steffen Windpassinger[1], Ralf Uptmoor[4], Wolfgang Friedt[1] & Rod J. Snowdon[1#]

ESM1: Supplementary Methods

Chilling tolerance evaluations

Chilling tolerance experiments were conducted at germination and during different seedling development stages. Unless specifically mentioned, all experiments were conducted using a randomised complete block design with a minimum of two and a maximum of nine replications per genotype. The different conditions used in the various experiments were designed to emulate potential scenarios for early spring weather conditions in temperate climatic zones. Optimal (control) conditions were chosen according to previously reported base temperatures for sorghum growth and development, considered to be 8.7°C for germination and 14.5°C for emergence and growth in a genotype-dependent manner (Brar et al. 1992; Payne, Batata & Rosenow 2003; Patanè et al. 2006; Fiedler et al. 2012).

Germination

Germination experiments were conducted under optimum (8h at 30°C, 16h at 20°C) and chilling conditions (13°C constant), as described by (Knoll et al. 2008). Two times fifty seeds were placed in two Jacobsen germination vessels. Percent germination (emergence of the

radical from the seed) was scored at 7 days after sowing for the optimal conditions and 15 days after sowing for the cold treatment.

Soil-based assay under controlled conditions

Emergence and seedling development in soil was evaluated in climate chambers fitted with 6 high pressure sodium SON-T AGRO 400 sodium lamps and 4 TL-D super 80 fluorescent lamps (Philips, Eindhoven, Netherlands), which provided 25,000 lux at 1m distance. Seeds were sown, 1.5 m from the light source, in 55 mm diameter and 55 mm deep QuickPot standard trays (Herkuplast Kubern GmbH, Germany) filled with Fruhstorfer soil (Hawita Gruppe GmbH, Vechta, Germany). Nine pots per genotype were sown with four seeds per pot. The experiment was run in parallel in two adjacent climate chambers.

Plants were grown under 11-hour days and 13-hour nights with two temperature regimes for optimum (25°C/20°C) and chilling (13°C/11°C) temperatures. The plants were watered every second day when necessary and humidity was maintained at 70% RH. The optimum condition trial was harvested 12 days after sowing, when most of the genotypes reached the two-leaf stage. The chilling trial was harvested in one variant after 1 month, and in a second variant two months after sowing, in each case at the two-leaf stage. Biomass development was determined by measuring seedling fresh weight (FW) and dry weight (DW), height, root length and root weight. Seedling root lengths were measured only for the chilling conditions, at one month after sowing.

Seedling survival was calculated from the total number of plants per genotype. Two types of survival rate were recorded; in the first case (survival 1) seedlings without strong wilting symptoms were scored as survivors, while in the second case (survival 2) plants were recorded as survivors only when they showed neither strong wilting nor significant chlorosis symptoms.

Semi-controlled soil-based assay

Greenhouse experiments were conducted in the winter of 2009/10 using the seeds produced in Italy. A fully climate-controlled greenhouse was used for the trial, with temperatures of ≤15°C (day) and 10°C (night) being used for the chilling treatment and 25°C/20°C for the optimal treatment. A 12h day/night regime with 70% RH was used in both variants. Seeds were planted in Fruhstorfer soil (Hawita Gruppe GmbH, Vechta, Germany) in 11cm*11cm*11cm pots and grown under 10,000 lux illumination from six sodium lamps. Plants were watered every second day.

Two chilling variants were conducted with different sowing temperatures. The first variant, designated chilling after emergence (CAE), was sown in six pots per genotype under optimal conditions (25/20°C) and transferred to the cold chamber when the emergence rate reached 50%. The second variant, designated chilling at sowing (CAS), was sown in two replications under the cold conditions and retained in the same chamber after emergence.

Emergence rate was counted every day, from the onset until cessation of emergence, which occurred between 5 and 29 days after sowing in the CAS experiment and between 3 and 8 days after sowing under optimum conditions. The optimal temperature trial was harvested 9 days after sowing and the chilling-stressed trial 40 days after sowing, at which time the majority of seedlings were at two-leaf stage or more. Seedling shoot and root length, shoot and root FW, shoot and root DW, and survival under chilling stress were determined at harvest.

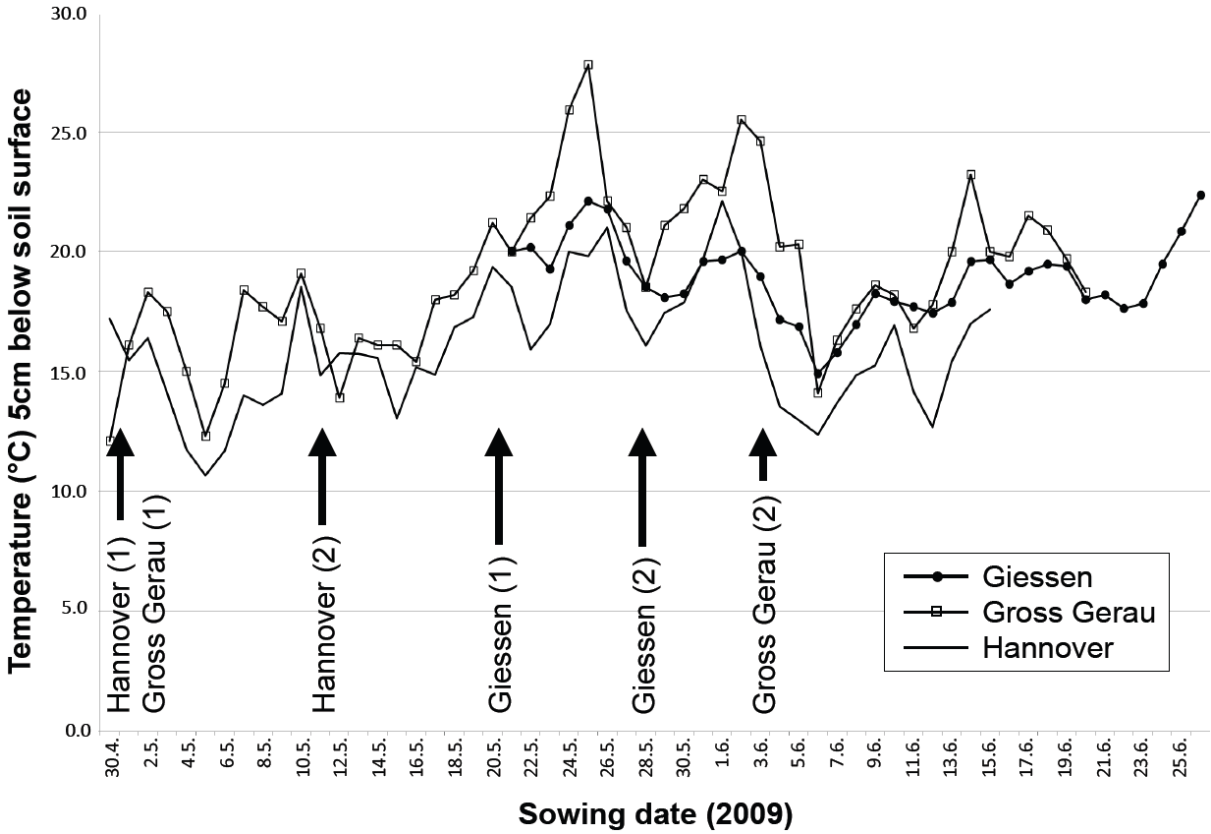
Selective phenotyping for validation of CAE survival

To validate chilling survival data from the 2010 soil-based assay, a total of 100 selected RILs, corresponding to the two extreme tails of the phenotypic distribution for the chilling survival

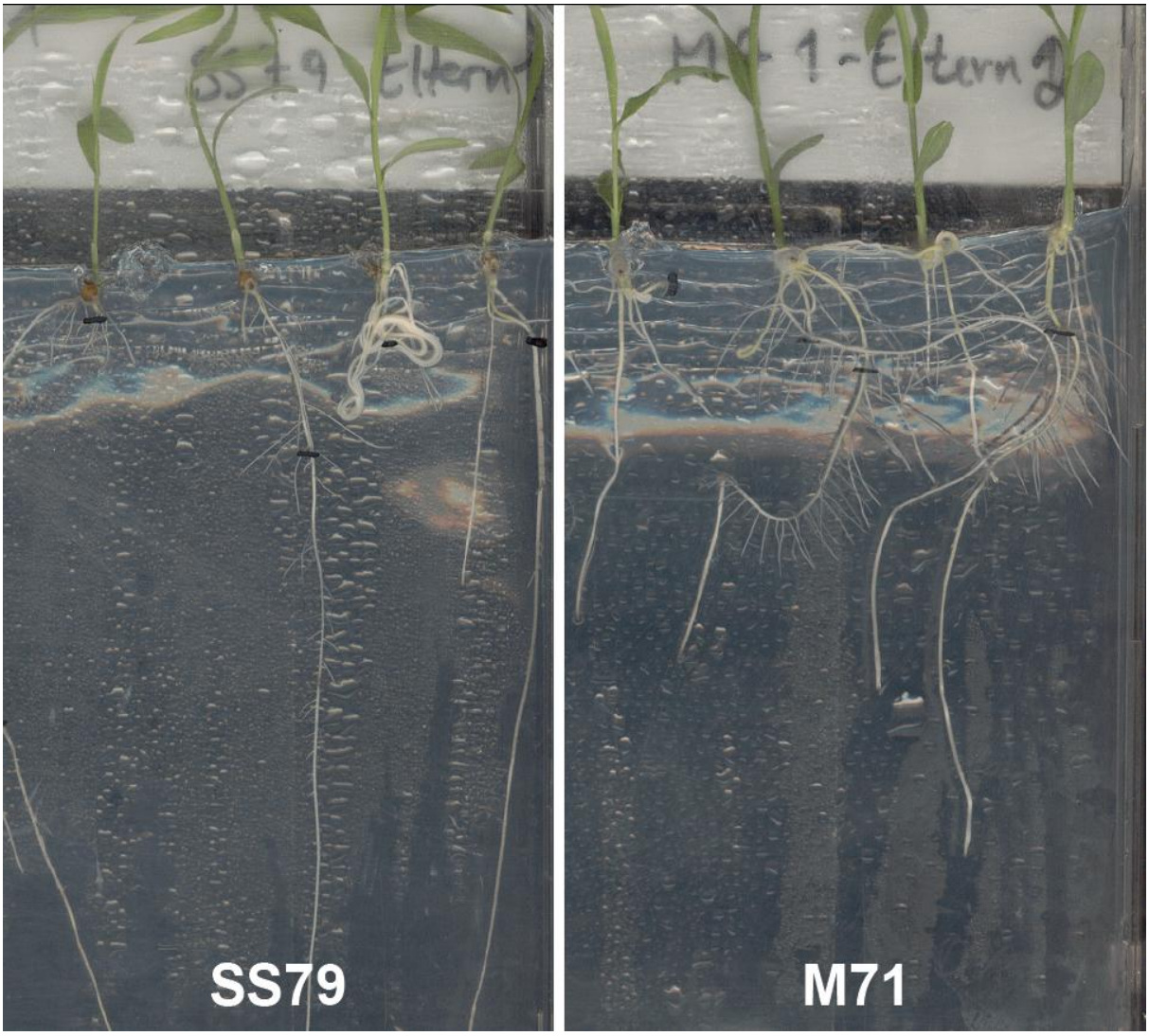
experiments, were sown in four replications on four consecutive days in the winter of 2010-11 and transferred to cold conditions (15°C/10°C) 7 days after sowing of the first replication. This meant that plants in the four replications were exposed to chilling stress from between 4 to 7 days over the four replications, allowing determination of the developmental time point at which chilling stress has the most serious influence on survival. The experimental design followed the CAE experiment described above. To exclude possible effects of soil-born pathogens, the experiments were conducted in sterilised sand media with irrigation being supplemented with 0.25 Murashige and Skoog medium (MS). Chilling survival rate along with root and shoot biomass were scored at the end of the experiment as described above.

ESM 2_Supplementary_Table_1.xlsx is attached to the thesis on a CD ROM.

ESM3_Supplementary_Figure1.pdf



ESM4_Supplementary_Figure2.pdf



ESM5_Supplementary Table 2a: Summary of main-effect QTL positions (Pos.), left flanking markers, 1-LOD support intervals, percent of explained phenotypic variance (R^2), standard effects (Std.) and additive effects (Add.) in the SS79 x M71 RIL population (n=178) from the 2009 chilling stress tolerance experiments.

Treatment	Trait	Chromosome	Pos. (cM)	Left flanking marker	LOD	Interval (cM)	R^2	Sum R^2	Std.	Add.
Cold	CCM 1 week	Sb02	16.6	Xtxp4	2.88	14.6-20.6	6.7	41.0	0.28	0.30
	CCM 2 weeks	Sb10b	14.8	Xtxp141	2.93	8.8-28.8	6.0	44.3	0.29	0.31
	Days to onset of germination	Sb02	10.6	E43M55_365	3.9	6.6-14.6	4.9	4.9	0.23	-0.30
	Percent emergence	Sb01	27.2	Xtxp357	4.31	25.2-29.2	10.1	62.3	-0.30	-0.30
	Percent germination	Sb01	27.2	Xtxp357	4.2	23-29.2	10.5	36.0	-0.31	-0.30
	Root DW at 1 MAS	Sb08	86	Xtxp292	2.51	84-96	5.0	17.8	0.27	0.49
	Root length soil 1 MAS	Sb02	2.6	Xtxp298	4.5	0-5	4.3	56.1	0.40	0.56
	Shoot FW at 1 MAS	Sb04b	26.4	Xtxp212	2.09	18.4-32.4	3.0	3.0	0.20	5.33
	Shoot DW at 1 MAS	Sb02	62.6	E32M52_355	4.46	58.6-64.6	2.9	36.5	0.15	0.39
	Shoot FW at 2 MAS	Sb01	47.2	E32M54_135	3.47	43.2-51.2	6	7.6	0.15	-4.32
	Shoot DW at 2 MAS	No QTL								
	Shoot length at 1 MAS	Sb03	46	E31M61_367	3.7	40-50	3.2	13.5	0.18	0.17
	Shoot length at 2 MAS	Sb06	8	E39M49_325	2.7	2-18	5.0	27.8	-0.34	-0.30
	Survival 1	Sb07	52	E44M60_142	3.66	44-58	2.7	18.0	0.16	5.10
	Survival 2	Sb01	111.2	Xtxp114	3.8	97.2-115.2	3.1	25.2	0.24	5.51
Optimum	CCM before stress	Sb01	25.2	E39M49_160	3.48	23.2-31.2	10.7	41.8	0.35	0.39

	Percent germination	Sb01	27.2	Xtxp357	7.64	25.2-29.6	22.5	88.3	-0.49	-7.60
	Shoot FW	Sb01	49.2	E32M60_364	3.72	45.2-51.2	6.3	6.3	-0.32	-0.80
	Shoot DW	Sb03	80	Xtxp114	8.66	74-82	3.4	34.4	-0.19	-0.10
	Shoot length	Sb02	30.6	E44M49_145	2.83	24.6-34.6	4.4	4.4	0.24	0.45
	Days to onset of germination	Sb01	115.2	E44M48_225	3.9	111.2-119.2	5.6	45.5	0.23	0.11
	Percent emergence	Sb01	27.2	Xtxp357	4.31	23.2-29.2	9.3	62.3	-0.26	-4.90
Seed lot	Seed color	Sb04b	30.2	Xtxp 212	3.9	28-36	3.3	31.9	-0.16	-0.10
	Thousand seed weight	Sb05	0	E31M61_195	3.24	0-2	4.6	15.6	0.25	1.44
Field	Perecent emergence	Sb04a	40	E32M58_115	2	30-48	5.6	14.7	0.28	5.97

ESM5_Supplementary Table 2b: Summary of pairwise epistatic QTL positions (between loci A and B) showing chromosome positions, left flanking markers, 1-LOD support intervals (cM), percent of explained phenotypic variance (R^2) and standard effects (Std.) in the SS79 x M71 RIL population (n=178) from the 2009 chilling stress tolerance experiments.

Treatment	Trait	Epistatic locus A				Epistatic locus B				R^2	Std.
		Chromo-some	Position (cM)	Left flanking marker	Interval (cM)	Chromo-some	Position (cM)	Left flanking marker	Interval (cM)		
Cold	CCM 1 week	2	16.6	Xtxp4	14.6-20.6	3	78	Xtxp114	74-80	4.40	0.2
						7b	6	Xcup57	0-14	3.30	-
	CCM 2 weeks	1b	1.7	E32M49_275	1.7-9.7	3	74	Stg9	68-80	6.20	0.2
						4a	0	E31M61_530	0-10	3.30	0.2
		3	24	E32M54_640	22-32	3	74	Stg9	68-80	3.00	-
		4a	0	E31M61_530	0-10	4a	28	Xtxp343	24-34	4.30	-
						10	14.8	Xtxp141	8.8-28.8	7.40	-
	CCM before stress	1	43.2	E32M47_298	39.2-47.2	1	25.2	E39M49_160	23.2-31.2	3.40	0.2
			57.2	E34M50_143	51.2-63.2	10	8.8	Xtxp141	8.8-20.8	7.90	-
			107.2	Dsb13	101.2-115.2	2	18.6	E43M62_390	2.6-18.6	3.70	-
		2	18.6	E43M62_390	14.6-20.6	3	30	E39M49_300	28-34	4.40	-
	Days to onset of germination	No epistatic QTL									
	Germination	1	51.2	E43M50_143	49.2-57.2	8	2	E43M44_357	0-26	3.10	-
		3	56	Stg17	48-62	9	74	E43M58_320	68-82	5.90	0.3
						4a	30.4	E32M55_297	26.4-38.4	3.50	0.2
	Percent emergence	1	27.2	Xtxp357	25.2-29.2	5	56	E31M59_210	46-68	4.30	-
						8	0	E43M49_357	0-10	3.90	-0.2
						9	74	E43M58_320	64-80	3.90	-
	Root DW soil 1 MAS	1	73.2	Xtxp299	55.2-89.2	1	1.2	E43M55_275	1.2-3.2	4.90	0.2
		2	12.6	Xtxp19	8.6-14.6	7b	14	Xcup57	8-14	3.60	-
	Root length soil 1	2	48.6	Xtxp50	46.6-52.6	7a	54	E43M58_349	40-60	5.40	-
			68.6	Xtxp286	64.6-70.6	2	80	E32M50_390	78-82	5.30	-

						7a	54	E43M58_349	40-60	4.40	0.2
	Seed colour	1	83.2	Xtxp61	71.2-97.2	9	50	E43M59_150	44-60	6.00	0.2
		4b	30.2	Xtxp212	28-36	9	4a	E34M50_315	0-12	6.40	-
		4a	10	E31M61_530	6-14	4b	4	Xcup5	0-14	11.00	0.3
	Shoot length at 2	1	27.2	xtxp357	23.2-31.2	4a	46	E35M49_230	18-50	4.10	0.2
		5	32	Xtxp14	18-34	6	72	E39M49_325	62-78	7.50	0.3
		6	58	Xtxp17	56-62	6	56	Xtxp265	52-60	4.10	-0.3
	Shoot DW at 1 MAS	1	3.2	E43M55_275	1.2-7.2	2	62.6	E32M52_355	58.6-64.6	7.50	0.3
						3	76	Stg9	72-78	4.60	-
						7	80	E31M59_202	66-84	5.70	-
			35.2	E35M49_260	31.2-39.2	8	92	Xtxp321	84-98	4.60	-
			71.2	Xtxp229	63.2-79.2	2	62.6	E32M52_355	58.6-64.6	7.50	-
						7	10	E45M57_120	0-22	3.70	-
			137.2	Xtxp316	123.2-139.2	1	71.2	Xtxp229	63.2-79.2	5.90	-
						4b	28.4	Xtxp212	24-32	3.60	0.2
	Shoot length at 1	1	31.2	E44M48_90	25.2-33.2	3	46	E31M61_367	40-50	4.90	0.2
						5	30	Xtxp15	28-34	4.80	0.2
	Shoot FW at 2 MAS	No epistatic QTL									
	Survival relaxed	1	11.2	E43M53_500	7.2-17.2	9	82	E43M51_210	74-84	5.60	0.3
			113.2	E44M48_225	97.2-119.2	2	8.6	E43M55_365	4.6-14.6	3.10	0.2
		4a	48	E35M49_230	44-50	7	52	E44M60_142	44-58	4.30	-
	Survival strict	1	125.2	Xtxp340	121.2-129.2	4a	42	E32M58_115	34-48	4.10	-
						9b	16	Xtxp289	10-16	3.40	-
			111.2	E44M48_220	97.2-115.2	8	44	E44M49_190	38-52	3.10	0.2
Optimal	Days to onset of germination	1	115.2	E44M48_225	111.2-119.2	7b	0	Xcup57	0-6	5.60	-0.2
						9	42	Xtxp230	38-48	11.60	0.3
			97.2	Dsb13	95.2-101.2	9	42	Xtxp230	38-48	11.60	0.3
		3	38	E45M59_256	34-40	5	38	E32M47_260	36-42	7.70	0.2
						10	4.8	Xtxp270	0.8-8.8	8.50	0.3
			66	Xtxp33	62-70	8	78	E39M49_210	70-80	5.10	0.3
		5	38	E32M47_260	36-42	10	4.8	Xtxp270	0.8-8.8	8.70	0.2

		6	30	Dsb37	26-48	8	78	E39M49_210	70-80	6.20	0.3
	Germination	1	27.2	357	25.2-29.2	8	0	E43M49_357	0-6	6.30	-
		3	26	E32M54_640	22-32	9b	16	Xtxp289	6-16	5.50	0.2
			52	Stg17	48-60	9	78	E43M51_210	71-82	7.00	0.3
	Percent emergence	1	27.2	Xtxp357	23.2-29.2	2	18.6	E43M62_392	16.6-20.6	3.70	-0.2
						9	0	E34M49_495	0-6	5.10	-
		1b	1.7	E32M49_275	1.7-9.7	5	58	E31M59_210	46-72	3.60	0.2
						9	0	E34M49_495	0-6	5.30	0.2
		2	54.6	E45M59_144	52.6-56.6	9b	0	Xtxp289	0-10	3.90	-
	Shoot FW	No epistatic QTL									
	Shoot DW	1	129.2	Xtxp248	125.2-131.2	3	80	xtxp114	74-82	12.30	-
		3	20	Xtxp31	18-22	9	26	E31M61_210	22-28	15.00	0.5
							68	E32M47_415	64-70	7.00	-
							84	E43M51_210	78-84	4.40	-
	Shoot length	No epistatic QTL									
Field	Field emergence	1	11.2	E43M53_500	7.2-19.2	4a	40	E32M58_115	30-48	5.10	0.2
		4a	40	E32M58_115	30-48	6	32	Dsb37	16-42	3.60	0.2
Seed lot	Thousand seed	5	0	E31M61_195	0-2	9	4b	Xcup2	38-54	5.10	0.3

ESM5_Supplementary Table 3a: Summary of main-effect QTL positions (Pos.), left flanking markers, 1-LOD support intervals, percent of explained phenotypic variance (R^2), standard effects (Std.) and additive effects (Add.) in the SS79 x M71 RIL population (n=177) from the 2010 chilling stress tolerance experiments. T_{50}: Time to reach 50% emergence; A_{max}:

Treatment	Trait	Chromosome	Pos.(cM)	Left flanking marker	LOD	Interval (cM)	R^2	Sum R^2	Std.	Add.
CAE	CCM 1 week	3	76	Stg 9	2.83	72-80	6.1	20.6	-0.774	-0.77
	CCM 2 weeks	No QTL								
	Root DW	2	28.6	Xtxp298	5.43	26.6-30.6	4.5	36.3	0.2	0.97
	Root FW	7	64	Xtxp227	3.75	56-74	4.6		-0.25	-5.24
	Root to shoot ratio	6	60.6	Xtxp17	4.54	58.6-62.6	3	10.4	0.31	0.13
	Root length soil	1	21.2	Xtxp248	4.57	19.2-23.2	5.5	40.3	-0.21	-0.87
		1b	1.7	E32M49_275	2.54	1.7-5.7	3.3		0.18	0.43
		2	48.6	Xtxp50	10.87	46.6-50.6	11.6		0.42	1.65
		4	12.4	Xtxp177	7.93	10.4-16.4	8.7		-0.35	-1.43
		6	70	E43M55_160	3.98	66-72	3.3		-0.18	-0.82
		8	38	E44M49_190	7.42	36-40	4		-0.19	-0.71
		9	46	Xcup2	4.29	42-50	4.9		-0.29	-0.83
			64	E34M50_195	2.64	60-66	3.4		0.23	0.62
		10	18	E32M58_195	9.6	16-20	5.2		0.21	0.99
	Shoot DW	6	74	E32M49_455	6.53	72-80	2.9	11.4	0.19	1.28
	Shoot FW	9	18	E31M61_210	6.8	12-22	3.6	3.6	-0.21	-4.28
	Shoot length	No QTL								
	Survival	1	127.2	Xtxp248	4.02	123-131	2.9	36.7	-0.21	-5.24
		6	28	Dsb37	3.17	22-34	3.9		0.24	6.69
	Root to shoot ratio	6	58	Xtxp17	4.54	56-60	3	15.4	0.31	0.13
CAS	Percent emergence	6	30	Dsb37	3.3	22-34	16.2	87.9	-0.64	-6.92
		9	50	E43M59_152	2.08	34-58	12.9		-0.36	-1.00
	Root length gel 15	6	24	Xtxp265	3.51	20-30	5.6	38.5	-0.23	-0.14

		8	84	E44M60_275	4.1	80-86	5.9		-0.25	-0.16
	Root length gel 10	3	22	E32M54_640	3.64	20-24	2.5	32.6	-0.16	-0.04
		6	70	E32M49_455	4.07	66-76	5.5		-0.21	-0.06
		8	84	E44M60_275	4.8	80-86	9.9		-0.31	-0.06
	Root elongation rate	3	64	Xtxp33	2.55	62-70	5.3	15.6	-0.29	-0.11
		6	24	Xtxp265	5.69	20-32	11.2		-0.35	-0.13
	Percent germination	6	6	E39M49_455	3.59	2-12	15	92.6	0.43	4.11
	Root FW	No QTL								
	Root DW	2	52.6	E45M59_144	5.34	51-56	4.0	15.1	0.21	0.63
	Root length soil	6	30	Dsb37	2.65	22-34	9.3	13.4	-0.34	-0.63
	Root to shoot ratio	1	107.2	Dsb13	6	101.2-111.2	6.4	18.8	-0.36	-0.22
	Shoot FW	1	75.2	Xtxp61	2.12	65.2-89.2	5	19.8	-0.23	-2.85
	Shoot DW	8	80	E44M60_275	4.2	76-84	5.6	5.6	-0.25	-0.36
	Shoot length	No QTL								
	Survival	1	51.2	E34M50_143	4.46	47.2-55.2	15.8	92.8	-0.32	-5.16
		6	28	Dsb37	9.07	24-34	16.5		0.35	11.99
		7b	14	Xcup57	4.43	8-14	21.8		-0.5	-9.10
	Days to onset of emergence	No QTL								
	Rate of emergence	3	16	E44M60_135	3.11	6-18	2.5	35.2	-0.19	-1.43
			42	Xtxp285	3.3	40-44	2.4		-0.19	-1.74
		6	24	Xtxp265	8.9	20-32	8.9		-0.28	-1.80
		8	78	E39M49_210	2.78	76-82	11		-0.39	-1.79
		9	50	E43M59_152	2.64	44-52	2.3		-0.13	-0.84
	A_{max}	2	24.6	Xtxp298	2.83	22.6-28.6	2.2	17	0.14	3.64
		6	28	Dsb37	3.38	20-34	7.2		-0.33	-6.48
	Emergence	No QTL								
	T_{50}	6	28	Dsb37	4.91	26-32	4.1	19.7	0.26	1.47
	Emergence day 2	No QTL								
Optimum	CCM before stress	1	139.2	Xtxp316	3.58	133.2-139.2	22.4	88.8	0.51	0.69

	3	40	Xtxp285	3.03	38-44	18.2		-0.54	-0.37
	6	48	E31M59_245	3.2	42-56	21.8		-0.75	-0.63
	8	4	E43M49_357	5.03	0-10	19.6		-0.7	-0.66
Percent emergence	9	64	E43M50_195	2.37	60-68	3.1	7.1	0.21	3.95
Root length gel 7	3	22	E32M54_640	9.3	20-24	9.3	15.6	-0.33	-1.33
	6	28	Dsb37	8.0	24-30	8		-0.29	-1.17
Percent germination	2	74.6	Xtxp286	4.03	63.6-76.6	7.9	41.5	0.31	6.33
	4a	22	E43M58_140	2	0-28	2.7		-0.16	-3.51
	6	56	Xtxp17	2.32	50-66	4.5		-0.22	-3.08
	7	54	E43M58_349	3.47	46-60	6.4		-0.25	-4.76
Root FW	No QTL								
Root DW	No QTL								
Root length soil	1	51.2	E34M50_143	3.01	49.2-55.2	11.5	23.3	-0.34	-0.54
Shoot DW	1	65	Xtxp279	2.05	47-79	3	20.1	-0.18	-0.49
	7	60	E43M58_349	2.46	52-70	2.7		-0.22	-0.69
Shoot FW	7	66	Xtxp227	5.43	56- 72	7	14	-0.31	-9.535
Shoot length	1	65.2	Xtxp279	2.21	56.2-73.2	4.5	4.5	-0.24	-0.50
Root to shoot ratio	6b	0	Sbagh04	2.93	0-12	2.9	4.5	0.22	0.08
Emergence day 1	2	56.6	Xtxp56	5.35	52.6-58.6	21.3	89.5	0.5	0.218
Rate of emergence	3	40	Xtxp285	2.59	34-42	2.9	7.2	-0.19	-3.41
A_{max}	No QTL								
Emergence	No QTL								
T50	3	58	Stg17	3.43	52-64	9.9	77.7	0.39	0.19
Emergence day 2	3	62	Xcup 38	3.25	54-64	7.2	30.4	0.37	0.29

ESM5_Supplementary Table 3b: Summary of pairwise epistatic QTL positions (between loci A and B) showing chromosome positions, left flanking markers, 1-LOD support intervals (cM), percent of explained phenotypic variance (R^2) and standard effects (Std.) in the SS79 x M71 RIL population (n=197) from the 2010 chilling stress tolerance experiments.

Treat-ment	Trait	Epistatic locus A				Epistatic locus B				R^2	Std.
		Chrom-osome	Position (cM)	Left flanking marker	Interval (cM)	Chrom-osome	Position (cM)	Left flanking marker	Interval (cM)		
CAE	CCM 1 week	3	76	Stg9	72-80	6	48	E31M59_245	36-54	2.8	-0.24
	CCM 2 weeks	No epistatic QTL									
	Root to shoot ratio	1	113.2	E44M48_225	107.2-117.2	6	70.0	E43M55_160	68-74	6.3	-0.41
						6	58.0	Xtxp17	56-60	10.3	0.61
	Root dry weight	1	55.2	E43M50_143	51.2-65.2	2	28.6	Xtxp298	24.6-30.6	5.9	0.27
						4	38.4	Xtxp51	30.4-42.4	3.6	0.24
			85.2	Xtxp61	85.2-87.2	9	80.0	E43M51_210	76-84	14.6	-0.63
			109.2	E44M48_220	107.2-115.2	6	58.0	Xtxo17	56-60	3.6	0.25
			131.2	Xtxp248	127.2-139.2	1	41.2	E32M47_298	39.2-45.2	4.6	0.26
						2	28.6	Xtxp298	24.6-30.6	5.1	0.23
	Root fresh weight	No epistatic QTL									
	Shoot fresh weight	No epistatic QTL									
	Shoot DW	1	85.2	Xtxp61	73-97	6	76	E32M49_455	72-80	6.2	0.37
	Shoot length soil	1	133.2	Xtxp316	131-139	9	52	E43M59_152	44-58	6.4	0.30

	Survival	1	127.2	Xtxp248	123.2-131.2	7	66.0	Xtxp227	60-72	5.4	0.52
		4b	26.4	Xtxp212	20.4-30.4	6	28.0	Dsb37	22-35	4.5	0.30
		4a	4	E31M61_530	0-10	6	28.0	Dsb37	22-34	3.1	0.23
		6	28	Dsb37	22-34	6	10.0	E39M49_325	4-20	7.2	-0.43
CAS	Root growth rate	1	133.2	Xtxp316	125.2-139.2	3	64.0	Xtxp33	62-70	4.2	-0.28
		4b	4.4	E44M49_348	0.4-22.4	6	24.0	Xtxp265	20-32	2.9	0.19
		5	82	E44M60_375	48-102	6	24.0	Xtxp265	20-32	3.5	-0.24
	Root length gel 10 DAS	1	25.2	E39M49_100	21.2-27.2	3	88.0	E32M53_150	84-88	4.1	-0.21
			47.2	E32M54_135	45.2-51.2	1	25.2	E39M49_100	21.2-27.2	7.5	0.33
		3	22	E32M54_640	20-24	3	68.0	Xtxp33	64-72	10.6	-0.51
						6	72.0	E32M49_455	66-76	5.5	0.25
	Root length gel107.2-117-2 15 DAS	5	60	E31M59_210	48-70	6	74.0	E32M49_455	66-78	2.8	-0.23
		8	84	E44M60_275	80-86	8	100.0	Xtxp321	90-108	3.0	-0.20
	Percent emergence	1	45.2	E32M54_135	53.2-69.2	5	10.0	E31M61_130	1-14	12.5	-0.55
		2	18.6	E43M62_390	14.6-22.3	6	6.0	E39M49_325	64-78	13.7	-0.89
			32.6	E44M49_145	30.6-36.6	7	52.0	E44M60_142	48-60	15.6	-0.70
			82.6	Xtxp7	82.6-78.6	6	8.6	E39M49_325	4.6-18.6	12.8	0.83
		3	44	E31M61_367	40-46	4b	2.4	E44M49_348	0.4-14.4	13.0	0.49
		4b	2.4	E44M49_348	0.4-14.4	9	50.0	E43M59_152	34-58	16.0	0.63
			36.4	Xtxp51	28.4-42.4	6	6.0	E39M49_325	64-78	12.8	0.78
	Percent germination	1	93.2	Xtxp61	81.2-109.2	2	80.6	xtxp7	68.6-82.6	19.9	0.61
						2	46.6	Xtxp50	44.6-48.6	21.6	-0.62

					3	36.0	E45M59_256	34-38	15.0	0.38
					8	96.0	Xtxp321	88-108	20.3	0.84
		115.2	E44M48_225	109.2-117.2	7	0.0	E44M57_120	0-6	33.1	0.92
					10	14.8	Pepc	2.8-24.8	14.9	0.53
		127.2	Xtxp248	125.2-129.2	7	0.0	E44M57_120	0-6	16.0	-0.66
					8	96.0	xtxp321	88-108	17.9	-0.99
	1b	9.7	E32M49_275	1.7-11.7	2	64.6	E31M59_126	54.6-66.6	23.0	-0.47
					9	68.0	E32M47_415	62-72	22.4	0.62
	2	18.6	E43M62_390	16.6-20.6	2	0.0	E32M50_390	0-4	18	-0.59
					6	6.0	E39M49_675	2-12	24.8	-0.76
		64.6	E31M59_126	54.6-66.6	8	38.0	E44M49_190	32-44	16.9	-0.43
Root dry weight	1	69.2	Xtxp229	63.2-75.2	2	52.6	E45M59_144	50.6-54.6	5.2	-0.28
					7b	0.0	Xcup57	0-8	4.1	0.23
Shoot FW	1	99.2	Dsb13	93-105	3	74	Stg9	66 - 78	5.4	0.29
	1	75.2	Xtxp61	65-89	2	66.6	E31M59_126	58.6 - 74.6	4.9	-0.27
	1	49.2	E44M60_364	45-51	4a	46	E35M49_230	40- 50	4.2	-0.25
Shoot DW	No epistatic QTL									
Shoot length soil	No espistatic QTL									
Root to shoot ratio	1	67.2	Xtxp229	63-69	6	68.0	E43M55_160	62-72	3.8	-0.26
		107.2	Dsb13	101.2-111.2	1	47.2	E32M54_135	45.2-49.2	4.6	0.26
					2	2.6	E32M50_390	0.6-6.6	7.4	0.33
		139.2	Xtxp316	137.2-139.2	1b	11.7	E32M49_275	5.7-11.7	4.9	-0.22
Root fresh weight	No epistatic QTL									
Root length	1	97.2	Dsb13	93.2-103.2	6	30.0	Dsb37	22-34	6.2	-0.27

	soil										
	Survival	1	51.2	E34M50_143	47.2-55.2	3	72.0	Xtxp33	68-82	14.3	0.41
						7	64.0	Xtxp227	58-78	15.8	0.45
						9b	16.0	Xtxp289	0-16	14.8	0.37
			121.2	Xtxp340	117.2-129.2	7	42.0	E44M60_142	38-46	17.5	0.48
						4	0.0	E44M49_348	0-4	16.6	-0.53
						6	30	Dsb37	24-34	14.1	0.43
		2	24.6	Xtxp298	22.6-32.6	9b	16.0	Xtxp289	0-16	15.6	0.43
		3	18	E44M60_135	16-20	6	2.0	E39M49_325	2-20	16.9	0.42
			72	Xtxp32	68-82	6	2.0	E39M49_325	2-20	18.4	-0.41
	Rate of emergence (S)	3	16	E44M60_135	6-18	6	24.0	Xtxp265	20-32	2.3	0.25
		4b	28	Xtxp177	22-36	8	78.0	E39M49_210	76-82	5.3	0.30
		6	24	Xtxp265	20-32	8	78.0	E39M49_210	76-82	2.4	0.30
						9	50.0	E43M59_152	44-52	3.3	0.18
	Amax	2	24.6	Xtxp298	22.6-28.6	6	8.0	E39M49_325	2-14	7.1	-0.23
		6	28	Dsb37	20-34	6	8.0	E39M49_325	2-14	7.1	0.42
	T50	6	28	Dsb37	20-34	6	6.0	E39M49_325	2-14	3.8	-0.33
Optimum	CCM before stress	1	139.2	Xtxp316	133.2-139.2	7b	0.0	Xcup57	0-8	15.3	0.44
	Root length gel 7 DAS	No epistatic QTL									
	Percent emergence	No epistatic QTL									
	Percent germination	2	74.6	Xtxp286	70.6-76.6	4a	22.0	E43M58_140	0-28	3.1	0.20
		6	56	Xtxp17	50-66	7	54.0	E43M58_349	46-60	3.1	-0.19
	Root to shoot ratio	No epistatic QTL									
	Root dry	No epistatic QTL									

weight										
Root fresh weight	No epistatic QTL									
Shoot FW	2	76.6	Xtxp7	74.6-82.6	2	28.6	Xtxp298	20.6-32.6	5.0	0.30
		28.6	Xtxp298	20.6-32.6	9b	16	Xtxp289	4- 16	4.8	0.25
Shoot DW	2	24.6	Xtxp298	20.6-28.6	7	60	E43M58_349	52- 70	3.4	-0.21
	4a	36	E32M58_115	10- 48	7	60	E43M58_349	52- 70	2.2	0.21
Root length soil	1	51.2	E34M50_143	49-55	3	34.0	E32M55_215	26-36	3.8	0.25
		139.2	Xtxp316	133.2-139.2	1	51.2	E34M50_143	49-55	4.1	0.21
Onset of emergence	1	113.2	E44M48_225	110-120	10b	10.0	Xtxp141	0-22	10.1	-0.50
	2	2.6	E32M50_390	2.6-12.6	4b	10.0	Xcup5	0-14	10.8	0.56
					10	6.0	E43M55_415	0-16	12.5	0.43
		26.6	Xtxp298	22.6-32.6	4a	16.0	E32M55_297	14-22	11.6	0.44
					9	32.0	E43M49_495	28-36	10.0	-0.42
		56.6	Xtxp56	52.6-58.6	5	46.0	E43M53_348	42-50	11.8	0.47
					8	76.0	E39M49_210	72-82	13.6	-0.44
	3	58	Stg17	50-64	4a	16.0	E32M55_297	14-22	14.5	-0.52
	5	46	E43M53_348	42-50	10	6.0	E43M55_415	0-16	14.8	-0.47
						10.0	Xtxp141	0-22	11.7	0.38
Rate of emergence (S)	1	89.2	Xtxp61	77.2-99.2	3	40.0	Xtxp285	34-38	3.0	-0.24
A_{max}	No epistatic QTL									
Uniformity	No epistatic QTL									
T_{50}	1	139.2	Xtxp316	135.2-139.2	2	56.6	Xtxp56	52.6-62.6	7.7	0.43
	2	2.6	E32M50_390	2.6-12.6	8	76.0	E39M49_210	72-78	7.5	0.49
	3	58	Stg17	52-64	4a	18.0	E43M58_140	14-22	11.9	-0.53

	4a	18	E43M58_140	14-22	10	24.8	E43M55_415	18.8-30.8	5.5	0.67
	4b	14	Xcup5	2-14	8	76.0	E39M49_210	72-78	6.5	-0.29
Cessation of emergence	3	62	Xcup38	54-64	6	66.0	Xtxp57	58-70	3.4	-0.25
					10	24.8	E43M55_415	18.8-28.8	3.4	-0.44
	6	66	Xtxp57	58-70	8	76.0	E39M49_210	72-78	4.0	-0.22
	10	24.8	E43M55_415	18.8-28.8	10	12.8	E32M58_195	16-22	4.1	-0.22

ESM5_Supplementary Table 4a: Summary of main-effect QTL positions (Pos.), left flanking markers, 1-LOD support intervals, percent of explained phenotypic variance (R^2), standard effects (Std.) and additive effects (Add.) in the SS79 x M71 RIL population (n=97) from the 2011 chilling stress validation experiments.

Trait	Chromosome	Position (cM)	Left flanking marker	LOD	Interval (cM)	R2	Sum R2	Std.	Add
Survival (chilling stress at 4-7 DAS)	2	6.6	E43M55_365	12.75	4.6-10.6	27.1	68.4	-0.42	-0.755
	4a	14.5	E32M58_115	9.12	10.5-24.5	12		-0.34	-6.502
		0	E35M49_230	6	0-2	7.4		0.23	-2.217
	4b	42	Xtxp51	11.2	40.4-42.4	5.8		0.19	1.701
	6	20	Xtxp97	11.35	18-26	16.7		-0.34	-5.196
	10	14.8	Pepc	10.97	12.8-16.8	7.3		0.17	3.706
	9b	0	Xtxp289	7.44	0-16	11.4		-0.26	-2.708
Survival (chilling stress at 6-7 DAS)	1	33.2	E44M48_90	4.64	31.2-37.2	8.2	61.6	-0.29	-5.643
		15.2	E44M60_395	3.43	11.2-19.2	15		0.22	1.983
	2	24.6	Xtxp298	9.28	22.6-26.6	10		-0.25	-1.072
	6	70	E43M55_160	4.43	66-72	5.9		-0.20	-10.707
	9	0	E34M50_315	3.66	0-14	15.5		0.29	4.072
Shoot FW	No QTL								
Shoot DW	8	10	E43M49_357	3.99	0- 14	6.1	21.8	0.26	1.208
Root FW	1	129.2	Xtxp248	6.48	127.2-139.2	15.9	47.3	0.38	7.897
	3	22	E32M54_640	12.06	20 - 24	14.3		-0.34	-12.163
Root DW	3	22	E32M54_640	5.72	20 - 24	19.0	57.7	-0.40	-2.364
	4a	50	E35M49_230	7.72	46 - 50	15.4		0.33	2.220
Root length	5	14	E32M47_198	3.66	12- 16	7.3	30.6	0.29	0.705
	6	66	Xtxp57	4.08	60 - 70	6.6		-0.31	-0.931
		34	Xcup12	6.79	18 - 30	18.7		0.46	1.442
Shoot length	6	34	Xcup12	5.36	30-38	9.6	18.5	0.32	1.404
	9b	0	Xtxp289	5.03	0- 10	5.9		-0.24	-0.807

ESM5_Supplementary Table 4b: Summary of pairwise epistatic QTL positions (between loci A and B) showing chromosome positions, left flanking markers, 1-LOD support intervals (cM), percent of explained phenotypic variance (R^2) and standard effects (Std.) in the SS79 x M71 RIL population (n=97) from the 2011 chilling stress validation experiments.

Trait	Locus A				Locus B				R^2	Std
	Chrom-osome	Position (cM)	Left flanking marker	Interval (cM)	Chrom-osome	Position (cM)	Left flanking marker	Interval (cM)		
Survival (chilling stress at 4-7 DAS	1	111.2	E44M48_220	107-113	1	49.2	E44M60_364	47-51	7.6	0.26
					3	18	E44M60_135	16-20	21	0.62
						38	E45M59_256	36-40	8.5	-0.46
		125.2	Xtxp340	123-127	1	35.2	E35M49_260	31-37	12.8	-0.32
					2	40.6	Xtxp304	38.6-42.6	21.2	0.42
					3	18	E44M60_135	16-20	7.3	-0.33
						38	E45M59_256	36-40	15.7	0.60
					10	16	Pepc	14-18	12.7	-0.37
Survival (chilling stress at 6-7 DAS)	1	1.2	E43M55_275	1-3	5	8	E31M61_130	0-28	13.3	0.39
		33.2	E44M48_90	31 -37	2	58.6	E32M52_355	54.6 - 62.6	17.8	0.44
					9	64	E34M50_195	56- 68	11.5	-0.32
		103.2	Dsb13	97-113	3	12	Xtxp9	2- 16	18.4	0.44
						88	E32M53_150	86-88	20.6	0.50
		131.2	Xtxp248	125.2-137.2	1	1.2	E43M55_275	1-3	7.9	0.25
Shoot DW	2	16.6	Xtxp4	10.6-20.6	8	10	E43M49_357	0-14	5.4	0.27
	4b	18.4	E43M49_170	4.4-22.4	4c	12	Xcup5	4-14	17.2	0.54
Root FW	1	15.2	Xtxp61	85.2-103.2	3	84	E32M53_150	80- 86	8.2	0.29
		113.2	E44M48_225	109-119	6	40	Xcup12	36-46	7.2	0.34
					8	76	E39M49_210	72 -78	10.2	0.36
		129.2	Xtxp248	127.2-139.2	1	41.2	E32M47_298	37.2-43.2	7.8	0.32
					4b	14.4	E44M49_348	12.4-18.4	9.1	0.30
Root DW	1	75.2	Xtxp61	63.2-83.2	8	108	Sb634	106- 108	8.2	0.24
		97.2	Dsb13	95-101	1	7.2	E43M53_500	1-11	8.8	0.24

				3	0	Xtxp9	0- 10	12.4	-0.28
				4b	28.4	Xtxp212	24.4-42.4	14.5	0.37
		128.2	Xtxp248	3-131	3	60	Stg17	54-74	10.5	0.35
Shoot length	No epistatic QTL									
Root length	No epistatic QTL									

ESM6_Supplementary Table 5. Positional and functional candidate genes located within main-effect and epistatic QTL hotspots for multi-trait chilling toelrance responses. The genes listed are those within the 1-LOD QTL support intervals that have annotations related to stress stimuli or meristem development (only for field emergence QTL on Sb04).

Chrom-osome	1-LOD support interval (cM)	Projected physical interval (bp)	Total number of genes	GO term	Genes in support interval with GO terms related to stress or cold response		
					Locus	Physical position	Annotation
Sb01	25-29	22459510..24002134	38	Abiotic stress stimuli	Sb01g020460.1	22905567..22909247	Ubiquitin-conjugating enzyme 5
					Sb01g020596.1	23338322..23338322	TRF-like 3
					Sb01g020800.1	23708432..23710576	TRF-like 8
					Sb01g020830.1	23829816..23831521	Peroxidase superfamily protein
					Sb01g020910.1	23937888..23942926	ATPase, AAA-type, CDC48 protein
					Sb01g020930.1	23992841..24002134	Phosphoglycerate kinase 1
	123-131	69117081..71117081	206	Abiotic stress stimuli	Sb01g046350.1	69556703..69558174	Heat shock transcription factor A6B
					Sb01g046460.1	69672935..69674464	Protein kinase superfamily protein
					Sb01g046570.1	69762208..69766448	P-loop containing nucleoside triphosphate hydrolases superfamily

		protein
Sb01g046630.1	69834462..69835417	PLAC8 family protein
Sb01g046650.1	69848452..69849494	Heavy metal transport/detoxification superfamily protein
Sb01g046700.1	69894337..69904092	Expansin 11
Sb01g046790.1	69965203..69969884	AMP-dependent synthetase and ligase family protein
Sb01g047040.1	70173877..70174405	Ubiquitin-conjugating enzyme 5
Sb01g047180.1	70400726..70402258	Ankyrin repeat family protein
Sb01g047360.1	70638155..70645194	Heat shock protein 60
Sb01g047430.1	70699947..70701467	Multidrug resistance-associated protein 5
Sb01g047450.1	70721339..70723371	myb domain protein 305
Sb01g047500.1	70754589..70756687	beta-amylase 1
Sb01g047540.1	70779897..70785003	carotenoid cleavage dioxygenase 1
Sb01g047930.1	71009451..71011231	sterol methyltransferase 2
Sb01g047940.1	71024191..71026244	beta HLH protein 93
Sb01g047970.1	71062804..71064161	Glutathione S-transferase F11

					Sb01g047980.1	71065168..71066341	Glutathione S-transferase phi 8
					Sb01g048000.1	71075448..71076863	Glutathione S-transferase phi 8
					Sb01g048010.1	71079932..71081320	Glutathione S-transferase F11
					Sb01g048040.1	71081594..71082583	Serine acetyltransferase 3;2
Sb04	Sb04a 30-48	49427601.. 51071003	81	Abiotic stress stimuli	Sb04g021150.1	49815964..49824634	DCD (Development and Cell Death) domain protein
					Sb04g021160.1	49831832.. 49833108	Integrase-type DNA-binding superfamily protein
					Sb04g021470.1	50522311.. 50531908	DNA/RNA helicase protein
					Sb04g021490.1	50539185.. 50542521	heat shock transcription factor A3
					Sb04g021560.1	50632948.. 50634152	soluble N-ethylmaleimide-sensitive factor adaptor protein 33
					Sb04g021670.1	50795397.. 50796321	RmlC-like cupins superfamily protein
					Sb04g021730.1	50853425.. 50855679	SOUL heme-binding family protein
				Meristem development	Sb04g021130.1	49780220.. 49787499	MEI2-like protein 1
					Sb04g021220.1	49989704.. 49993123	indeterminate(ID)-domain 2
					Sb04g021440.1	50488431.. 50491220	indeterminate(ID)-domain 2

					Sb04g021500.1	50557581.. 50563466	Protein kinase superfamily protein
					Sb04g021620.1	50725428.. 50725760	protein binding
Sb06	22-32	50720303.. 52760303	198	Abiotic stress stimuli	Sb06g021550.1	50786625..50789950	Peroxidase superfamily protein
					Sb06g021570.1	50798753.. 50800781	Damaged DNA binding;DNA-directed DNA polymerases
					Sb06g021650.1	50858886.. 50864113	Protein phosphatase 2C family protein
					Sb06g021680.1	50874889.. 50877327	DNA glycosylase superfamily protein
					Sb06g021730.1	50936345.. 50936937	Heavy metal transport/detoxification superfamily protein
					Sb06g021890.1	51113671.. 51115583	VIER F-box proteine 1
					Sb06g021900.1	51123276.. 51124793	UDP-Glycosyltransferase superfamily protein
					Sb06g021950.1	51187252.. 51202414	MUTL protein homolog 1
					Sb06g022025.1	51270656.. 51274741	Sensitivity to red light reduced protein (SRR1)
					Sb06g022060.1	51302335.. 51304922	Glutaredoxin family protein
					Sb06g022140.1	51362892.. 51367323	General transcription factor II H2
					Sb06g022180.1	51402429.. 51408913	Dicer-like 4

Sb06g022190.1	51416580.. 51419285	Dicer-like 4
Sb06g022200.1	51425577.. 51427714	Dicer-like 4
Sb06g022230.1	51467178.. 51475592	Ammonium transporter 1;1
Sb06g022280.1	51501642.. 51503654	seed gene 1
Sb06g022310.1	51528848.. 51534737	RING/U-box superfamily protein
Sb06g022340.1	51566002.. 51569605	Breast cancer associated RING 1
Sb06g022540.1	51748059.. 51751722	NB-ARC domain-containing disease resistance protein
Sb06g022550.1	51753675.. 51757601	NB-ARC domain-containing disease resistance protein
Sb06g022660.1	51880841.. 51882874	Myb domain protein 15
Sb06g022720.1	51948477.. 51949094	SAUR-like auxin-responsive protein family
Sb06g022810.1	52026442.. 52029170	Auxin response factor 16
Sb06g022840.1	52066384.. 52068417	Plasma membrane intrinsic protein 2
Sb06g022870.1	52085298.. 52085916	Scorpion toxin-like knottin superfamily protein
Sb06g022940.1	52151795.. 52153389	UDP-glucosyl transferase 73D1

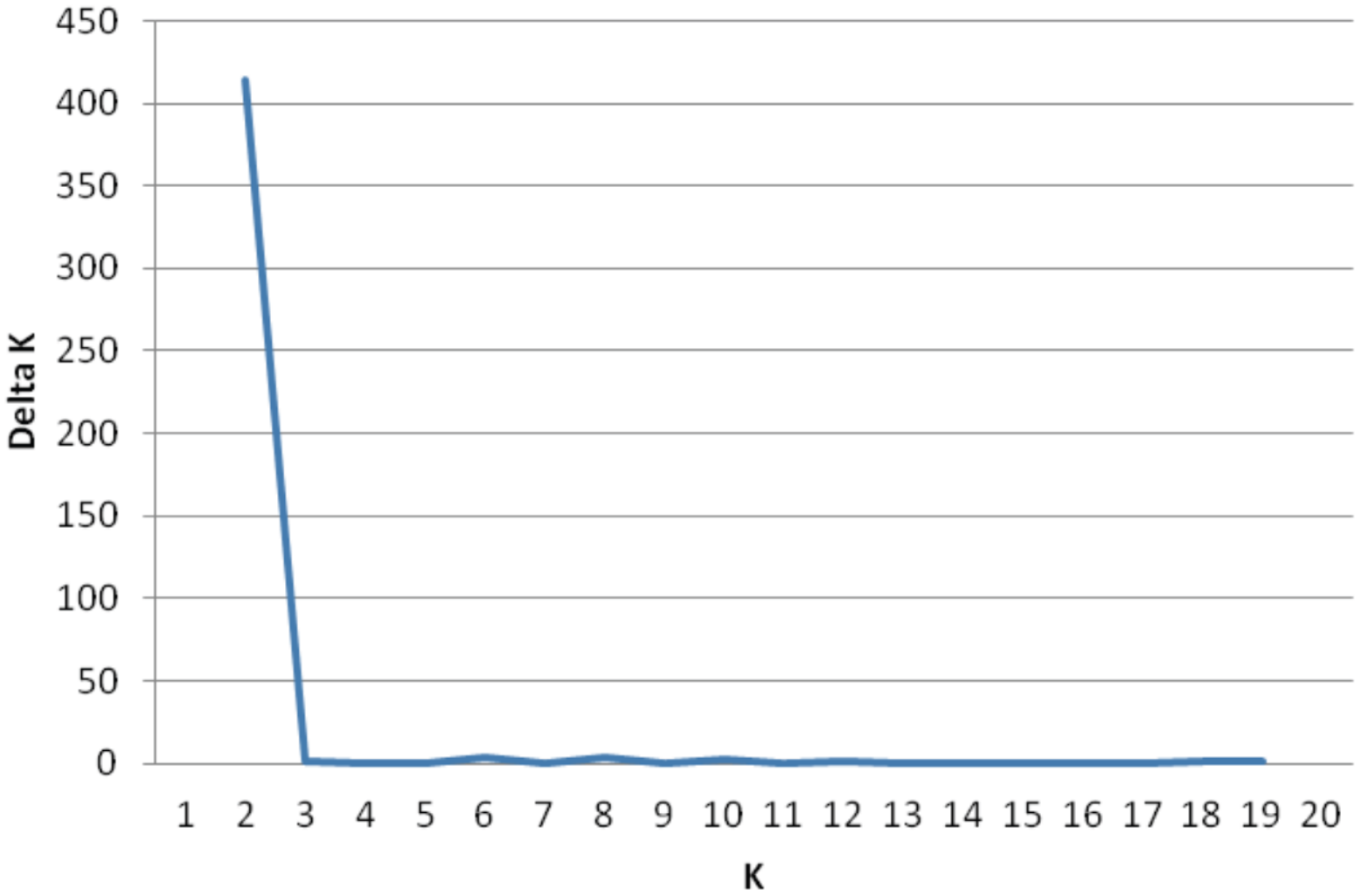
Sb06g022950.1	52155901.. 52157412	UDP-glucosyl transferase 73D1
Sb06g022960.1	52174367.. 52175234	response regulator 4
Sb06g023010.1	52224706.. 52228830	YELLOW STRIPE like 7
Sb06g023020.1	52229245.. 52232403	YELLOW STRIPE like 7
Sb06g023100.1	52282617.. 52287578	serine carboxypeptidase-like 35
Sb06g023120.1	52295810.. 52298217	nitrilase 4
Sb06g023125.1	52302443.. 52311255	Major facilitator superfamily protein
Sb06g023130.1	52340767.. 52341345	TCP family transcription factor
Sb06g023230.1	52405379.. 52407825	Major facilitator superfamily protein
Sb06g023240.1	52408577.. 52409875	beta-tonoplast intrinsic protein
Sb06g023290.1	52437225.. 52439764	Hus1 like gene
Sb06g023330.1	52405379.. 52407825	Major facilitator superfamily protein

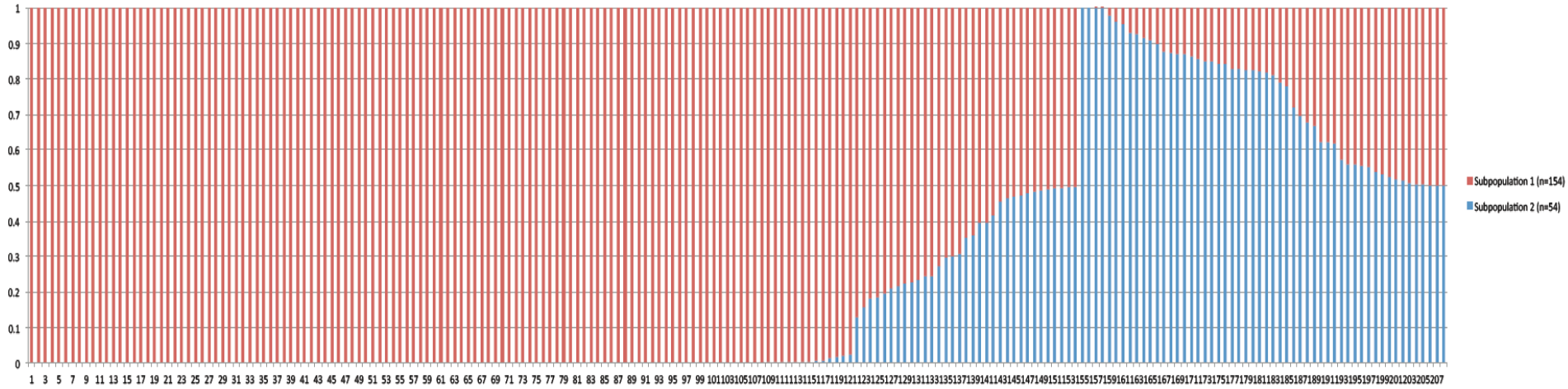
ESM7_Supplementary_Figure_3.pdf is on the attached CD ROM.

Appendix II: Electronic supplementary materials from Bekele et al. (2013b)

ESM1: An excel file containing the SNP information is attached to the thesis (CDROM).

Bekele_PBJ_ESM2_ Structure.pdf.




1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79 81 83 85 87 89 91 93 95 97 99 101103105107109111131151171191211231251271291311331351371391411431451471491511531551571591611631651671691711731751771791811831851871891911931951971992012032052 07
Subpopulation 1 (n=154)
Subpopulation 2 (n=54)

Appendix III: Supplementary materials from Chapter 4

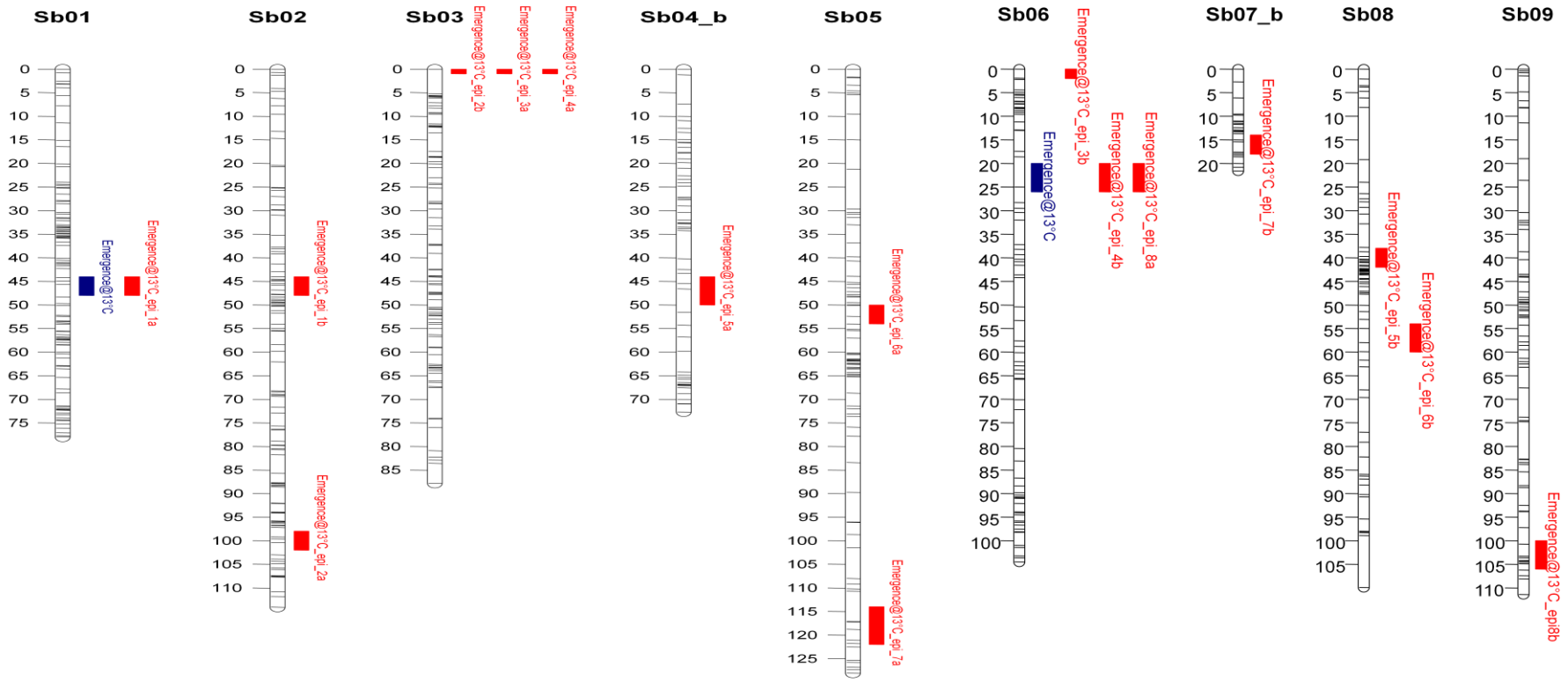


Figure 1 Epistatic (red bars) and main effect QTL (blue bars) detected in 92 RILs under chilling stress conditions (13°C/10°C) from the 2010 trial (Bekele et al. 2013a). The distance of the bars shows the 1 LOD support interval of the QTL.

Figure 2 Comparison of sorghum cold tolerance QTL studies and rice cold germination studies. Colour-coding for sorghum studies: Bekele et al. 2013, Burow et al. 2011, Fiedler et al. 2012, Knoll et al. 2008; Rice study: Li et al. 2013. (TAG 126: 2313-2322)

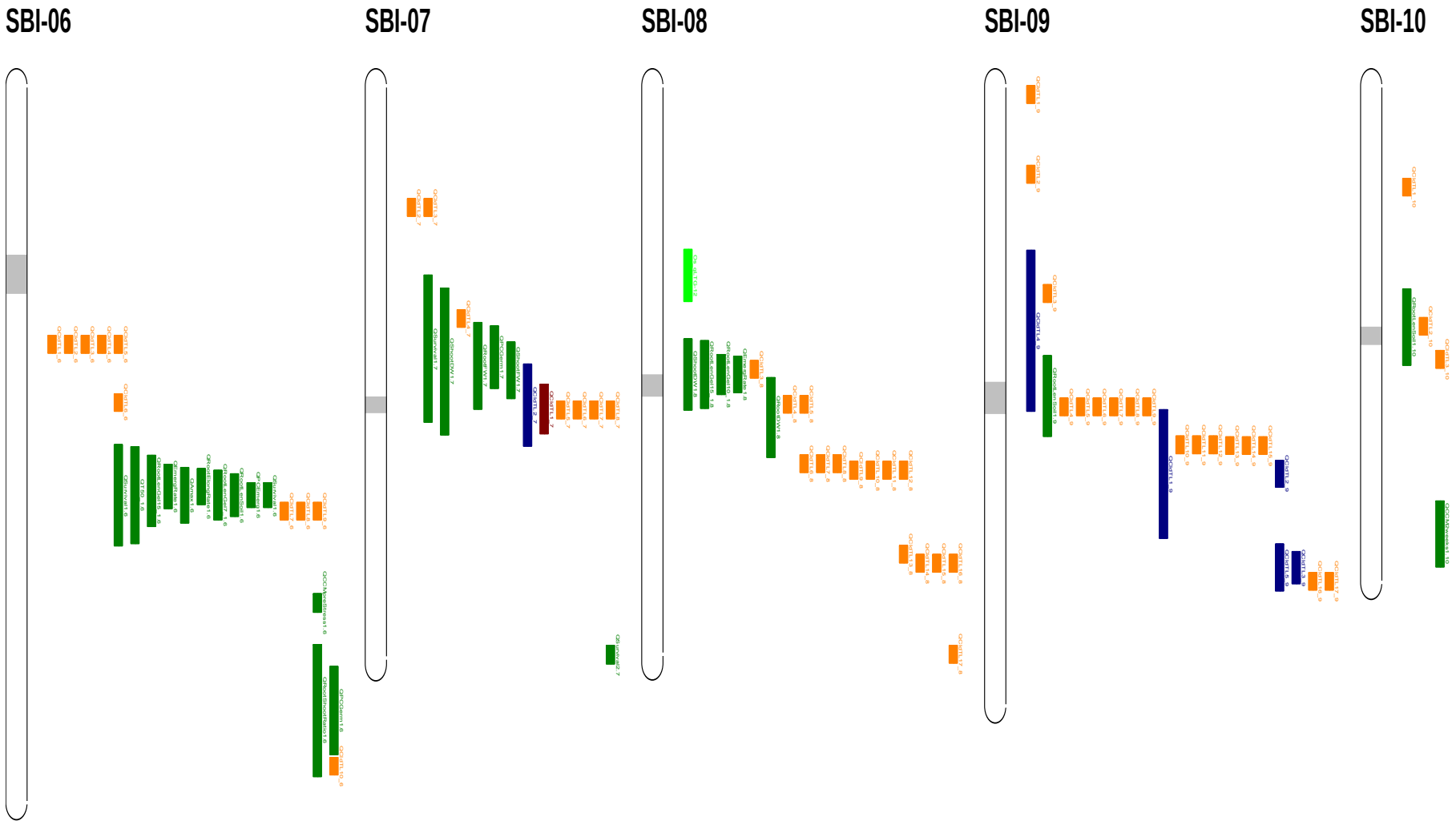


Figure 2 Comparison of sorghum cold tolerance QTL studies and rice cold germination studies. Colour-coding for sorghum studies: Bekele et al. 2013, Burow et al. 2011, Fiedler et al. 2012, Knoll et al. 2008; Rice study: Li et al. 2013. (TAG 126: 2313-2322)

Table 1 Detailed QTL tables for field drought data

Main effect QTL

Trait	Treatment	Chromosome	Position	Left flanking marker	Support interval	Partial R^2	Std
Plant height	Irrigated	Sb02	20	UGSS_01910	16-22	13.3	0.26
		Sb07	14	UGSDII_11250	12-16	54	-0.75
		Sb10	106	UGSDI_52464	98-110	15.7	-0.35
	Non-irrigated	Sb01	52	UGSDII_01356	50-54	8.4	0.16
		Sb02	12	UGST_00150	10-16	17.3	0.21
		Sb06	34	UGSS_04764	30-40	7.0	-0.15
		Sb07	18	UGSDI_30811	14-20	46.8	-0.44
CCM	Irrigated	Sb08	56	UGSDII_14035	54-58	39.8	0.36
	Non-irrigated	Sb10	44	UGSDI_43530	38-46	8.4	-0.28
Brix	Irrigated	Sb04	0	UGSDI_19293	0-2	10.2	0.41
		Sb06	4	UGSDI_26793	2-6	20.7	-0.41
		Sb07	14	UGSDII_11250	12-16	19.5	-0.37
		Sb10	24	UGSS_00165	16-28	10.5	0.28
	Non-irrigated	Sb01	10	UGSDI_00189	6-12	10.6	1.37
		Sb06	10	UGSDII_09504	8-12	20.6	-1.53
		Sb06	72	UGSDII_10183	70-78	12.6	-0.55

Epistatic QTL

		Locus 1		Locus 2			
Trait	Treatment	Chr.	Position	Chr.	Pos.	Part R2	Std
Plant height	Irrigated	Sb06	76	Sb10	106	18.4	-0.35

		Locus 1		Locus 2			
Trait	**Treatment**	**Chromosome**	**Position**	**Chr.**	**Pos.**	**Part R^2**	**Std**
Plant height	Non-irrigated	Sb01	0	Sb06	54	10.51	0.16
			12	Sb04	20	15.1	-0.20
				Sb10	8	8.9	0.15
			52	Sb04	42	26.8	-0.30
		Sb02	78	Sb04	4	20.0	0.29
		Sb04	4	Sb07	18	13.5	-0.17
			20	Sb06	68	14.1	0.20
			42	Sb07	18	7.2	-0.14
		Sb05	10	Sb06	54	17.5	-0.26
		Sb06	34	Sb10	40	10.9	0.18
			54	Sb06	68	19.6	0.34
			96	Sb07	16	8.5	0.14
CCM	Irrigated	Sb01	14	Sb07	26	36.4	0.33
			50	Sb02	0	8.3	0.13
			64	Sb04	30	44.5	-0.55
			76			16.5	0.24
		Sb02	42	Sb02	60	41.1	-0.54
				Sb07	0	22.5	0.24
			60	Sb06	0	39.9	0.35
			90	Sb05	6	43.5	-0.50
		Sb03	4	Sb03	60	14.4	0.26
			84	Sb05	74	21.2	-0.17
		Sb05	16	Sb09	40	22.5	-0.24

		Locus 1		Locus 2			
Trait	**Treatment**	**Chromosome**	**Position**	**Chr.**	**Pos.**	**Part R^2**	**Std**
CCM	Irrigated	Sb05	74	Sb06	12	51.1	-0.53
		Sb06	12	Sb10	100	21.5	-0.30
		Sb07	0	Sb07	27	23.7	0.24
			28	Sb08	72	28.3	0.26
		Sb10	100	Sb10	130	38.9	-0.43
	Non-irrigated	Sb03	24	Sb03	58	5.7	0.34
		Sb03	58	Sb10	70	6.9	0.37

Table 2 Details of the emergence QTL from the 2010 emergence data (Bekele et al. 2013a) and using the genetic linkage map developed using SNP array data (Bekele et al. 2013b) based on 92 RILs from SM population**.**

Main Effect QTL

Trait	**Treatment**	**Chr.**	**Pos.**	**Left flanking marker**	**Supp. inter.**	**Part R^2**	**Std**
Emergence	Cold	Sb01	46	UGSDI_02682	44-48	8.4	-0.23
		Sb06	24	UGSDII_09855	20-26	47.9	-0.67
	Optimum	Sb02	0	UGSDII_01729	0-2	12.5	0.36

Epistatic QTL

		Locus 1		**Locus 2**			
Trait	**Treatment**	**Chromosome**	**Position**	**Chr.**	**Pos.**	**Part R2**	**Std**
Emergence	Cold	Sb01	46	Sb02	26	13.3	-0.28
		Sb02	100	Sb03	0	19.4	0.33
		Sb03	0	Sb06	0	10.4	0.27
					24	24.4	0.40
		Sb04	46	Sb08	42	16.3	0.35
		Sb05	52		56	15.0	-0.38
			118	Sb07	16	23.5	-0.39
		Sb06	24	Sb09	104	16.0	0.31

Supplementary materials and methods used

1. Description of the field drought experiment (GG2009)

The field experiment was conducted in Gross Gerau experimental station as described by (Shiringani et al. 2010). The only difference was the plot size ($6m^2$) and the two treatments were two times irrigation and no irrigation (stress) during the growing season. The following table shows the statistical summary of the experiment.

Table 3 Statistical summary of the 2009 field trial in Gross Gerau

Trait	Parental lines		RIL population (N=188)			Significance		
	SS79	M71	Mean	Min	Max	Irrigation	Genotype	Gen * Irr
Plant height (cm)	217.5	113.3	151.0	75.0	243.0	**	**	**
	292.5	137.5	190.6	75.8	340.8			
CCM	22.1	27.6	28.6	13.7	63.6	**	**	**
	33.3	31.8	30.6	13.5	118.0			
Brix	18.8	14.6	17.3	9.5	36.7	**	**	**
	16.8	15.9	15.8	9.6	20.6			

Shaded part shows the trait performance under stress condtions ** significant at P=0.01

2. Programs and softwares used for QTL analysis, association mapping and genomic selection

2.1. QTL mapping was undertaken as in Bekele et al. 2013a using Plabqtl program at higher LOD (5.0) threshold for co factor selection.

2.2. Association mapping was undertaken using the TASSEL software used in the Bekele et al. 2013b. GLM using population structure and marker from Bekele 2013b and phenotype data from Fiedler et al. 2012 at temperatures (12.3°C) and (19.9°C) were used.

2.3. The R based rr-BLUP program (Edelman 2011; Edelman et al. 2012) was used for genomic selection on 92 RILs from SM population described in (Bekele et al. 2013b).

List of Abbreviations

ANOVA	Analysis of variance
bp	base pairs
CCM	Chlorophyll content meter
CIMMYT	Centro Internacional de Mejoramiento de Maíz y Trigo
CIM	Composite interval mapping
cM	centi Morgans
DH	Doubled haploid
DNA	Deoxyribonucleic acid
FAO	Food and agriculture organization of the United Nations
g	gram
Gb	Giga base pairs
GLM	General linear model
GS	Genomic selection
h^2	heritability
ha	hectar
ICRISAT	International Crop Research Institute for Semi-arid and Arid Tropics
Indel	Insertion and deletion polymorphisms
JLAM	Joint linkage and association mapping
k	Kilo
LD	Linkage disequilibrum
LOD	Logarithm of odds
Mb	Mega base pairs
MAS	Marker assisted selection
NAM	Nested association mapping
PAV	Presence-absence variation
QTL	Quantitative trait locus
RIL	Recombinant inbred line
ROS	Reactive oxygen species
SNP	Single nucleotide polymorphism

Declaration

I declare that the dissertation here submitted is entirely my own work, written without any illegitimate help by any third party and solely with materials as indicated in the dissertation. I have indicated in the text where I have used texts from already published sources, either word for word or in substance, and where I have made statements based on oral information given to me. At all times during the investigations carried out by me and described in the dissertation, I have followed the principles of good scientific practice as defined in the "Statutes of the Justus Liebig University Gießen for the Safeguarding of Good Scientific Practice".

Acknowledgments

I would like to thank my “Doktorvater”, Professor Wolfgang Friedt, for the tremendous kindness, support and advice he has offered me ever since I joined his Lab as a Masters student. You have gone extra miles to finance my studies, arrange visa and related issues. You believed in my capabilities, respected my opinions and encouraged me to thrive further. As a result, in addition to my PhD degree, I have gained experience in teaching, student supervision, networking and research management. I am always grateful for these opportunities.

I am grateful to my second supervisor Prof. Matthias Frisch who agreed to evaluate my work. Your office was always open whenever I needed advice on statistical genetics aspects of my work.

My deepest gratitude goes to my immediate supervisor and the current head of the Institute of Plant Breeding (Prof. Rod Snowdon). I am grateful to your patience, understanding and close supervision. I have learned a lot from your insights and experience in scientific writing. Your “out of the box” thinking was a perfect complement to my ambitious plans. Your success in fund raising has made this expensive project a reality. I am grateful to you and your family for several dinners, memory games with the children and for the German translation of the summary of my thesis (Konny).

I would like to thank all members of the Institute of Plant Breeding. Almost everyone has contributed towards the completion of my thesis work. I would like to thank Dr. Amukelani Shiringani for providing me with the mapping population used in this study and the associated data, for our heated debates, productive discussions, lots of fun and a lasting friendship. Swetlana Renner and Nelly Weis are the best technical assistances one can ever imagine. We had fun in the green house, on the field or in the laboratory. Technical assistants at IPZ-IFZ were all helpful in green house and climate chamber phenotypings and of course driving to and from experimental stations. Stavros was always there to help me set up my customized root and drought phenotyping methods and had always time for my thousands of questions with regards to lab, green house equipments and consumables. Liane Renno, Annete Planck, Birgit Keiner and Anja Pöltl, gave hands at some point in my series of experiments. Ich bedanke mich Schatzis! Petra Kretschmer and Burkhard Lather helped in green house management. Benjamin Wittkop was a nice office mate and did much of the driving and we discussed extensively on field trials and biochemical analysis. Thanks! Christian Obermeier, I thank you very much for our after work discussions on sequencing, bioinformatics, mapping etc. I still credit you for the Linux course, though at the end I was beating your MySQL codes and spoiling the keyboard with refreshment drinks. Several bachelor and masters students did part of their work attached to my project and I am grateful to their contributions: Steffen Windpassinger, Daniel Schanubelt, Kai Voss-Fels, David Wiese, Tim Bernhard, Helge Fluss, Esther Shaab and Ariane Malinowski.

The technical assistants and the other staffs of the experimental stations have done a tiresome job in an excellent way. My gratitude goes to Gross Gerau station head Mario Tolksdorf and the friendly staff who made me feel like part of their group. Markus Kolmer has organized

Phytotrone and rain out shelter experiments including a demanding drought trial, which involved root washing in Rauischolzhausen. His proactive approach and meticulous work was impressive. Giessen field station staffs headed by Malte Luh and Martin Seim have undertaken all the experiments with high efficiency. Thank you!

I would like to thank the project partners of the Bio-Energie 2021 project for the successful collaboration project: Karin Fiedler and Ralf Uptmoor from the University of Hannover; Arndt Zaccharias, Silke Wieckhorst and Millena Ouzunova from KWS Saat AG. I would like to thank the funding agencies (BMBF, DAAD and BMLE) that supported our work and local and international travels.

I would like to extend my gratitude to local and international research groups collaborating with our group. I would like to thank Professor Hai-Chun Jing at the Chinese Academy of Sciences, Institute of Botany, for the provision of pre-publication whole genome sequence data and ongoing collaboration on denovo sequence assembly. I would like to thank the friendly and helpful collaboration with Dr. Korbinian Schneeberger and Vipul Patel from Max-Planck institute for plant breeding research. They have contributed SHORE map detected SNP data and I hope we will soon finalize the QTL-sequencing project initiated some time ago. I am grateful for the wonderful discussions of the Australian group who paid us a visit recently: Professor Ian Godwin, Professor David Jordan, Dr. Emma Mace and Dr. Bradley Campbell from the University of Queensland. Collaboration on fine mapping and mutation screening will definitely further strengthen our collaboration.

Many friends from Unterhof (Reemt, Henry, Michael, Silvia, Britta, Moustafa, Isabellaa, Malgosiata, Hamza and Rajka) made my stay in Giessen unforgettable. The Ethiopian students' community in Giessen (Martha & Darah, Betre, Mengistu, Markos, Engudi, Abraham, Dawit, Hewan, Anteneh, Kidist and many more) was a family abroad. Thank you for organizing several events and Ethiopian holiday celebrations!

My comrade-in-plant sciences & more (Dr. Blen Beyene) and my three little sisters (Essete, Melhik and Hiyaw) are the sources of my inspiration. I thank you very much for your prayers, love and support. My father (Abi), who raised my sisters and me as a single parent for most of our lives, paid lots of sacrifice. I have no words to describe your love, patience and trust. I know that this means a lot to you – making you proud and happy is the driver of my life. I am always thankful to you and of course to the Almighty God.