Setaria viridis and Setaria italica, model genetic systems for the Panicoid grasses

Pinghua Li1 and Thomas P. Brutnell1,2,*
1 Boyce Thompson Institute, Cornell University, Ithaca, NY 14853, USA
2 Department of Plant Biology, Cornell University, Ithaca, NY 14853, USA
* To whom correspondence should be addressed. E-mail: tpb8@cornell.edu

Received 18 January 2011; Revised 1 March 2011; Accepted 9 March 2011

Abstract

Setaria italica and its wild ancestor Setaria viridis are diploid C4 grasses with small genomes of ~515 Mb. Both species have attributes that make them attractive as model systems. Setaria italica is a grain crop widely grown in Northern China and India that is closely related to the major food and feed crops maize and sorghum. A large collection of S. italica accessions are available and thus opportunities exist for association mapping and allele mining for novel variants that will have direct application in agriculture. Setaria viridis is the weedy relative of S. italica with many attributes suitable for genetic analyses including a small stature, rapid life cycle, and prolific seed production. Setaria sp. are morphologically similar to most of the Panicoideae grasses, including major biofuel feedstocks, switchgrass (Panicum virgatum) and Miscanthus (Miscanthus giganteus). They are broadly distributed geographically and occupy diverse ecological niches. The cross-compatibility of S. italica and S. viridis also suggests that gene flow is likely between wild and domesticated accessions. In addition to serving as excellent models for C4 photosynthesis, these grasses provide novel opportunities to study abiotic stress tolerance and as models for bioenergy feedstocks.

Key words: Abiotic stress tolerance, biofuels, C4 photosynthesis, domestication, Setaria italica, Setaria viridis.

Introduction

The Panicoideae subfamily is one of the most agronomically important grass groups. It includes the most economically important C4 cereal in the USA, maize, as well as several important grains grown throughout the world including sorghum, pearl millet, and foxtail millet. It also includes the major biofuel, sugarcane, as well as emerging bioenergy feedstocks Miscanthus and switchgrass. Both Setaria italica (common name: foxtail millet) and Setaria viridis (common name: green foxtail) are members of the tribe Paniceae that is the sister to the Andropogoneae. The Andropogoneae tribe includes maize, sorghum, Miscanthus, and sugarcane (Fig. 1). Of plants with sequenced genomes, S. viridis and S. italica (http://www.phytozome.net/foxtailmillet.php) are the closest relatives of the biofuel feedstock switchgrass and the invasive weed guinea grass (Panicum maximum). As a genetic model, S. viridis has many desirable traits, including small size (10–15 cm), a short life cycle (6–9 weeks depending on photoperiod conditions), and prolific seed production (~13,000 seeds per plant). Plants are self-compatible and can be transformed by Agrobacterium-mediated methods (Brutnell et al., 2010). Setaria italica is a cereal crop with excellent drought tolerance and an extensive germplasm collection, providing opportunities to study the domestication process and to mine for novel allelic variation. Collectively, these traits suggest that S. viridis and S. italica are ideally suited for studies of C4 evolution, comparative grass genomics, and use as models for biofuel feedstocks (Doust et al., 2009; Brutnell et al., 2010).

As whole-genome sequencing projects progress at an impressive rate (Schnable et al., 2009; Vogel et al., 2010), it is likely that new model organisms will emerge to address specific challenges in agriculture, for example biotic and abiotic stress tolerance. Currently, Arabidopsis thaliana serves
as the community standard to investigate many developmental, disease, and abiotic stress responses. However, monocots and eudicots diverged 127–175 million years ago (Bell et al., 2010; Smith et al., 2010), providing ample opportunity for divergence of gene structure and function. In this review, the potential of *S. italica* and *S. viridis* as models to investigate the mechanisms driving domestication and to understanding photosynthetic differentiation using a systems biology approach is discussed. In addition, the *Setaria* sp. provides novel opportunities to address one of the most pressing agronomic challenges in world agriculture today, namely abiotic stress tolerance.

**Distribution of *Setaria italica* and *Setaria viridis***

*Setaria italica* is one of the oldest crops in the world; cultivation probably began 5900 BP in Gansu Province, Northwestern China (Barton et al., 2009). Currently, *S. italica* is a minor food crop in China and India, and in the USA, Canada, and Australia is grown primarily for fodder or bird seed (Baltensperger, 2002; Diao, 2005, 2007; Doust et al., 2009). Some cultivars of *S. italica* are extremely drought tolerant, making them major crops in the arid regions in the North of China where annual rainfall amounts can be as low as 100–400 mm (Diao, 2005, 2007).

Although *S. viridis* originated around latitude 45°N in Eurasia, it has since spread to a wide range of habitats in temperate, tropical, and subtropical regions (Dekker, 2003). Its successful history of invading, colonizing, and adapting to the local environment, together with simple growth requirements and large seed yield probably contributed to its domestication, and it is largely accepted that *S. viridis* is the wild ancestor of cultivated *S. italica* (Li et al., 1945; de Wet, 1975). This hypothesis has been supported by isozyme and phylogenetic analyses using ribosomal and chloroplast sequence markers as well as through chromosomal fluorescent *in situ* hybridization (Wang et al., 1995; Benabdelmoula et al., 2001; Doust and Kellogg, 2002). The remarkable adaptation of *S. viridis* to a host of environments has contributed to its dubious honour as one of the worst weeds in world agriculture, successfully competing with many cereal crops (Dekker, 2003).

**Genetic diversity of *Setaria italica* and *Setaria viridis***

Surveys of *S. italica* and *S. viridis* have revealed a germplasm rich with phenotypic variation. A comparison of 1535 *S. italica* accessions collected from 26 countries at the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in India identified large variations in plant height, flowering time, inflorescence architecture, and seed morphology (Reddy et al., 2006). A sample of a few *S. viridis* accessions also reveals large variations in plant height, flowering time, and inflorescence structure (Fig. 2). In contrast to the many accessions of *S. italica* available (Doust et al., 2009), few accessions of *S. viridis* are available in the USA (GRIN: http://www.ars-grin.gov/cgi-bin/npgs/html/taxon.pl?430573). Clearly, additional collections of *S. viridis* from around the world would be of great value in exploring the diversity of this important plant group.

Despite the phenotypic diversity, the genetic diversity underlying this variation does not appear to be as great. A comparison of isozyme variation among 164 accessions of *S. viridis* suggests low genetic diversity within *S. viridis* populations, similar to estimates of genetic diversity in other self-pollinating weeds. However, this diversity is
distributed in discrete populations, suggesting strong local adaptation and regional differentiation (Wang et al., 1995). More recent amplified fragment length polymorphism (AFLP) and single nucleotide polymorphism (SNP) marker analyses support these findings of low genetic variation within accessions and high variation between populations (d’Ennequin et al., 2000). The silent nucleotide variation of S. viridis (θa=0.0059) is more than double that found in S. italica (θa=0.0027), suggesting a loss of diversity during domestication (Wang et al., 2010a). This drop in diversity is similar to what has been observed in other self-pollinating grasses such as wild (θa=0.0077–0.0095) and domesticated rice (θa=0.0024) (Zhu et al., 2007). A pairwise SNP analysis across nine loci indicated an increase of linkage disequilibrium corresponding to a loss of genetic diversity in S. italica relative to S. viridis and identified a potential selective sweep at one locus during the domestication process (Wang et al., 2010a). However, Fst estimates between S. viridis and S. italica varied from 0.025 to 0.41 among the nine genes sampled (Wang et al., 2010a). This broad range in Fst estimates suggests that a deeper sampling of both genes and accessions is necessary to reveal the degree of population differentiation between S. viridis and S. italica. It is also worth noting that the close physical proximity in which S. viridis and S. italica are often found, together with a lack of apparent hybridization barriers, probably permits cross-hybridization and gene flow between weedy and domesticated relatives. The estimated rate of outcrossing varies from 0.3% to as high as 4%, suggesting ample opportunity for gene flow between Setaria species, particularly from crop to weedy species (Wang et al., 2010a). One caveat to all current diversity studies is the relatively small sample size of S. italica and S. viridis accessions taken from the putative centre of origin in Gansu province. Increasing the richness of the current germplasm collections to include multiple accessions closer to the centre of origin and domestication of S. italica (Barton et al., 2009) and sampling a much broader segment of the genome should facilitate studies of genetic diversity and gene flow between populations.

Model genetic systems for understanding the domestication process

Most of the morphological and physiological differences between S. italica and S. viridis correspond to effects of human selection or the ‘domestication syndrome’ (Doyley et al., 2006). Compared with its wild ancestor, many S. italica accessions have reduced vegetative branching, an enlarged inflorescence, synchrony of flowering, loss of seed dormancy, and loss of disarticulation. Much of our understanding of domestication in Panicoideae grasses has come from comparisons of maize with its wild progenitor, teosinte (Doyley and Stce, 1991; Lukens and Doyley, 1999). However, due to morphological and physiological differences between maize and other grasses, it is likely that different suites of genes have been targeted during the domestication of maize relative to other Panicoide crops. For example, tb1, a major domestication quantitative trait locus (QTL) underlying tillering in maize, does not appear to be a major control point of branching variation differentiating S. italica from S. viridis (Doust et al., 2004). Although several loci appear to regulate inflorescence architecture in Setaria (Doust et al., 2005), a comparative analysis of maize and Setaria QTLs that regulate branch morphology will be necessary to reveal both common and unique targets of domestication. Thus, the study of domestication traits in S. italica and S. viridis, which are morphologically most similar to the majority of Panicoideae grasses including many of the emerging bioenergy feedstocks, could identify novel targets of selection and suggest loci for marker-assisted selection.

Flowering time is an important trait in cereals, which directly influences grain yield (Jung and Muller, 2009) and is the result of breeding selection for adaptation to specific photoperiod and temperature environments (Colasanti and Coneva, 2009). Comparative studies between Arabidopsis, rice, and Brachypodium have revealed that many aspects of photoperiod control of flowering are conserved between eudicot and monocot lineages (Higgins et al., 2010); however, homologous genes may be under very different regulation. For instance, the CONSTANS gene acts as a positive regulator of flowering under long days in A. thaliana, but in rice, the homologous gene functions to repress flowering under long days (Hayama and Coupland, 2004). Furthermore, the vernalization pathways are significantly different between eudicot and monocot lineages (Higgins et al., 2010). For instance, FLC acts as a central regulator of vernalization in A. thaliana, but no functional homologue is present in monocot lineages (Cockram et al., 2007). Conversely, in wheat, a zinc-figure CCT domain protein ZCCT1 (also known as VRN2) acts as a dominant repressor of flowering, but a clear homologue is absent in Arabidopsis (Yan et al., 2004). Since flowering time is such an important trait related to grain yield, it is perhaps not surprising that significant variation exists in the mechanisms to alter the timing of flowering in crop species. Thus, even in closely related crops such as wheat and rice, there is significant divergence in the genes that influence flowering time (Cockram et al., 2007). Study of the genes regulating flowering in pairs of cultivated and wild species may be more insightful in understanding the evolution of this important domestication trait rather than making wide cross-species comparisons. As S. viridis has one of the widest global distributions of any weed (Dekker, 2003) and S. italica is cultivated under both long-day and short-day conditions, there is probably substantial genetic variation in genes regulating photoperiod- and vernalization-dependent flowering time. In summary, comparative studies between S. viridis and S. italica will probably contribute much to our understanding of how flowering time can be manipulated during the domestication process.

An alternative approach to identify domestication genes is through whole-genome scans, which can be used to compare DNA polymorphisms between wild and domesticated species

Setaria as a model for C4 photosynthesis | 3033
to define the regions with reduced variation consistent with selection (Storz, 2005). Next-generation (NextGen) sequencing techniques provide more power for this approach (Wang et al., 2010b), which can efficiently map genome-wide SNP variation (McNally et al., 2009; Huang et al., 2010). These techniques could be applied to characterize *S. italica* and *S. viridis* germplasm to identify selective sweeps, and determine markers for future breeding efforts. In addition to targeting genes for food production, it should also be possible to exploit these data for engineering bioenergy feedstocks, for example by identifying genes from *Setaria* sp. that repress inflorescence formation and delay senescence and thus increase biomass production. Because *S. italica* and *S. viridis* are close relatives to bioenergy feedstocks such as switchgrass, it is likely that these findings will be relatively easy to translate from *Setaria* into less tractable genetic systems such as switchgrass, *Miscanthus*, or sugarcane.

### Model genetic systems for understanding C₄ photosynthesis

Despite the economic value of C₄ grasses, the regulatory networks that initiate and maintain C₄ photosynthesis are largely unknown (Brutnell et al., 2010). This gap in our understanding is a major roadblock in current attempts to bioengineer C₄ traits into C₃ crops such as rice (Mitchell and Sheehy 2007; Sheehy et al., 2007; Zhu et al., 2010). The recent sequencing of the sorghum and maize genomes (Paterson et al., 2009; Schnable et al., 2009) coupled with a detailed analysis of the maize leaf transcriptome (Li et al., 2010) provides new opportunities to investigate gene networks underlying C₄ photosynthesis. However, due to the long life cycle, large plant size, and lack of efficient transformation systems for these crop plants, neither sorghum nor maize is ideal for the genetic dissection of C₄ traits. Thus, it is time to explore *Setaria* sp. as potential models for understanding C₄ photosynthetic development.

C₄ photosynthesis evolved at least 17 times independently within the grass family (Duvall et al., 2003; Christin et al., 2009b). Thus, by using a comparative genomics approach, it should be possible to identify both common and unique signatures of C₄ evolution. For instance, a comparison of several C₄ enzymes, including phosphoenolpyruvate carboxylase (PEPC), phosphoenolpyruvate carboxykinase (PCK), and NADP-dependent malic enzyme (NADP-ME), among multiple lineages has revealed signatures of strong positive selection during C₄ evolution (Christin et al., 2007, 2009a, c). Although much less is known about common cis-regulatory elements, it is likely that both shared and unique regulatory elements were recruited to drive high levels of gene expression in either bundle sheath or mesophyll cells (Yanagisawa and Sheen, 1998). One obvious difference among the grasses is in the choice of enzyme used to decarboxylate the major C₄ acid in the bundle sheath cells. There are three primary decarboxylase enzymes that were recruited to C₄ photosynthesis, NADP-ME, NAD-dependent malic enzyme (NAD-ME), and PCK (Leegood, 2002). In this regard, the Paniceae clade is particularly interesting as all three subtypes exist. For instance, switchgrass uses NAD-ME for decarboxylation, guinea grass uses PCK, and *Setaria* uses NADP-ME. Thus, by comparing the cis-regulatory regions of these decarboxylases among these three lineages, it may be possible to discover common features that typify changes that accompany the cell-specific control of gene expression in C₄ photosynthesis.

As mentioned before, one emerging technology that promises to reshape the analysis of plant genomes is NextGen sequencing (Lister et al., 2009; Wang et al., 2010b). The recently released whole-genome sequences of two Andropogonea C₄ grasses, sorghum and maize, has already provided some insight into the mechanisms of gene duplication that are likely to drive C₄ differentiation (Wang et al., 2009). Furthermore, with the decreasing cost and increasing throughput from emerging sequencing technologies, it will soon be possible to deeply profile gene expression across numerous tissues and developmental time points. For instance, a comparative analysis of *S. italica* and *S. viridis* genomes in the Paniceae with maize and sorghum in the Andropogoneae will allow the comparison of two independently derived NADP-ME subtypes of C₄. Additional sequencing of both C₃ and C₄ grasses distributed throughout the grass group will enable a much broader examination of the genome dynamics that has accompanied C₄ evolution. This will include comparisons of not only coding regions, cis-regulatory elements, and splice sites, but also large-scale and small-scale genome rearrangements and methylation profiles. Furthermore, as additional grass genomes are sequenced, comparative genomic approaches will become increasingly powerful. Thus, over the next few years, ultra high throughput sequencing technologies promise to provide an unprecedented opportunity to explore the evolution and diversification of C₄ photosynthesis in the grasses. As one of the smallest, fastest growing grasses with an efficient transformation protocol, *S. viridis* is well positioned as a model for functional genomic dissection of C₄ photosynthesis.

### Model grasses for abiotic stress tolerance study

The range of environments in which *Setaria* sp. are found suggests that the germplasm may also be a rich source of genetic variation for genes controlling abiotic stress tolerance. For example, some *S. italica* cultivars are remarkably drought tolerant; for 1 g of dry biomass, maize requires 470 g of water, wheat 510 g, and *S. italica* a modest 257 g (Diao, 2005, 2007). This high water use efficiency (WUE) has helped establish *S. italica* as a major crop in the semi-arid and arid regions in north China (e.g. Gansu, Shanxi, and Neimenggu) and India. Several morphological and physiological adaptations are associated with increased WUE in *S. italica*, including small leaf area, thickening of the cell walls, and ability to form a dense root system (Li, 1997). To identify genes underlying these traits, Zhang and colleagues...
and Lata and colleagues examined differential gene expression among drought-tolerant and susceptible accessions of *Setaria* following stress treatment and identified hundreds of significant changes in expression that corresponded to genes involved in metabolism, protein synthesis, and signaling (Zhang et al., 2007; Lata et al., 2010). Although a wide range of salt (Sreenivasulu et al., 2004; Jayaraman et al., 2008; Puranik et al., 2010) and cold (PL, unpublished results) tolerance is evident in the *Setaria* germplasm, the mechanistic basis of this variation is unknown.

Although abiotic stress tolerance and photosynthesis are often studied in isolation, it is important to consider the effects of interaction of abiotic stress on photosynthetic activities. In particular, plants that maintain high photosynthetic rates under abiotically stressed conditions (e.g. high salt, drought, and poor soil nutrition) will be particularly important as more marginal lands are incorporated into agriculture, especially for bioenergy feedstocks. The sequencing of the *S. italica* and *S. viridis* genomes and the characterization of transcriptomics changes associated with abiotic stress treatments will provide new and exciting opportunities to investigate these networks.

**Genomics resource development**

Given the relatively small size of the *S. italica* and *S. viridis* genomes, surprisingly little progress has been made in establishing genetic tools for these plants. To date, only two QTL mapping populations have been published (Devos et al., 1998; Wang et al., 1998). One population was generated as an intraspecific cross between two *S. italica* accessions (Longgu 25×Pagoda Flower Green). The other was the result of an interspecific cross between *S. italica* and *S. viridis* accessions (*S. italica* B100×*S. viridis* A10). An additional recombinant inbred line population has also recently been generated between *S. italica* accession Yugu1 and *S. viridis* accession A10 (Doust et al., 2009). Although many QTL studies have been performed in *Setaria*, few are publicly accessible. The Gramene database indicates that at least 65 QTLs have been defined in screens of the B100×A10 population (http://www.phytozome.net/foxtailmillet_intro.html). As recently discussed, efforts are now underway to generate a high density map of the Yugu1×A10 population which should help to better define intervals in QTL studies (Doust et al., 2009).

The recent whole-genome sequencing of *S. italica* (JGI, http://www.phytozome.net/foxtailmillet.php) and ongoing sequence analysis of *S. viridis* will greatly accelerate the genetic analysis of these two species as well as facilitate comparative analyses within the grasses. By combining genetics maps, QTL studies, and genome sequencing, it will soon be possible to identify candidate genes that underlie QTLs in *Setaria*. In addition, the establishment of transient and stable *Agrobacterium*-mediated transformation in *S. viridis* (Brutnell et al., 2010), virus-induced gene silencing in *S. italica* (Doust et al., 2009), and mutagenized populations of *S. viridis* will provide valuable tools for the molecular characterization of gene function in *Setaria*. Thus, through the development of new genomics tools and technologies, it will soon be possible to add *Setaria* to the growing list of model organisms in the grasses that will aid in comparative genomic approaches to characterize the world’s most important feed, fuel, and bioenergy feedstocks.

**Conclusion**

The weedy *S. viridis* and domesticated *S. italica* species promise to serve as excellent model systems for the study of *C₄* photosynthesis, artificial selection, abiotic stress tolerance, and biomass production in the Panicoid grasses. Importantly, the completion of the *Setaria* genome sequence as well as the development of efficient transformation methods will rapidly advance these studies. What is needed now is a detailed anatomical, physiological, and biochemical exploration of these two species to provide the framework for the molecular genetic dissection of agronomically and economically important traits in the grasses.

**Acknowledgements**

We thank Hugues Barbier and Lin Wang for critical reading of the manuscript. This work was supported by grant IOS-0701736 from the National Science Foundation to TB.

**References**


