

# Genetic diversity and origin of North American green foxtail [*Setaria viridis* (L.) Beauv.] accessions

Stephan Schröder · Bochra A. Bahri · Douglas M. Eudy ·  
Daniel J. Layton · Elizabeth A. Kellogg · Katrien M. Devos

Received: 16 July 2015 / Accepted: 11 January 2016 / Published online: 25 January 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** *Setaria viridis* (L.) P. Beauv. and its domesticated form, *S. italica* (L.) P. Beauv., have been developed over the past few years as model systems for C<sub>4</sub> photosynthesis and for the analysis of bioenergy traits. *S. viridis* is native to Eurasia, but is now a ubiquitous weed. An analysis of the population structure of a set of 232 *S. viridis* lines, mostly from North America but also comprising some accessions from around the world, using 11 SSR markers, showed that *S. viridis* populations in the US largely separate by latitude and/or climatic zone. *S. viridis* populations from the Northern US and Canada (north of 44°N) group with accessions from Western Europe, while populations in the Mid and Southern US

predominantly group with accessions from Turkey and Iran. We hypothesize that *S. viridis* in the US was most likely introduced from Europe, and that introductions were competitive only in regions that had climatic conditions that were similar to those in the regions of origins. This hypothesis is supported by the fact that Canadian *S. viridis* lines were fast cycling and undersized when grown in the Mid-Western and Southern US compared to their morphology in their native environment. A comparison of the population structure obtained with 11 SSR markers and ~40,000 single nucleotide polymorphisms (SNPs) in a common set of *S. viridis* germplasm showed that both methods essentially yielded the same groupings, although admixture was identified at a higher frequency in the SNP analysis. Small numbers of SSR markers can thus

**Electronic supplementary material** The online version of this article (doi:10.1007/s10722-016-0363-6) contains supplementary material, which is available to authorized users.

S. Schröder · B. A. Bahri · D. M. Eudy ·  
K. M. Devos (✉)  
Department of Crop and Soil Sciences, Institute of Plant  
Breeding, Genetics and Genomics, University of Georgia,  
Athens, GA 30602, USA  
e-mail: kdevos@uga.edu

S. Schröder · B. A. Bahri · K. M. Devos  
Department of Plant Biology, University of Georgia,  
Athens, GA 30602, USA

**Present Address:**  
S. Schröder  
Department of Plant Sciences, North Dakota State University,  
Fargo, ND 58108, USA

B. A. Bahri  
Department of Plant Protection and Postharvest Diseases,  
National Agronomic Institute of Tunisia, 1082 Tunis,  
Tunisia

D. J. Layton · E. A. Kellogg  
Department of Biology, University of Missouri-St. Louis,  
St. Louis, MO 63121, USA

**Present Address:**  
E. A. Kellogg  
Donald Danforth Plant Science Center, St. Louis, MO 63132,  
USA

be used effectively to discern the population structure in this inbreeding species.

**Keywords** Adaptation · Genetic diversity · Green foxtail [*Setaria viridis* (L.) Beauv.] · Population structure · Simple sequence repeat

## Introduction

*Setaria viridis* (L.) Beauv., green foxtail, belongs to the family *Poaceae*, subfamily *Panicoideae*, tribe *Paniceae*. It is the wild progenitor of foxtail millet [*S. italica* (L.) Beauv.], a crop cultivated for food mainly in China, India and Russia, for birdseed in Europe and for hay and silage in the United States. *S. viridis* can cross freely with the cultivated *S. italica*, and this has been exploited to transfer traits from *S. viridis* to *S. italica* including herbicide resistance (Naciri et al. 1992; Wang et al. 1996). Naciri et al. (1992) reported that only few backcrosses were needed to eliminate weediness following an inter-specific cross between *S. viridis* and *S. italica*, making trait introgression from the weed to the cultivated species a viable strategy. *S. viridis* may also be a source of novel stress tolerance genes for the genetic improvement of *S. italica* (Qie et al. 2014).

*Setaria viridis* is an annual diploid species thought to be native to Eurasia (Darmency 2005) but now a ubiquitous weed in temperate regions throughout the world (Invasive Species Compendium: [www.cabi.org/isc/](http://www.cabi.org/isc/)). In North America, *S. viridis* was first reported in Montreal, Canada in 1821 (Douglas et al. 1985). Most likely it was introduced as a contaminant of crop seed and in the ballast of ships. It remained a relatively minor weed in Canada until at least the 1930s (Manson 1932), but by 1948, the species was widespread throughout Manitoba, Alberta and Saskatchewan (Groh and Frankton 1949). In the US, *S. viridis* has been present since at least 1900 and has greatly increased in abundance over the past 100 years (Forcella and Harvey 1983). *S. viridis* is considered one of the most successful plants in colonizing disturbed habitats. It is typically found in agricultural fields, on road sides and along railroad tracks, on ditch banks and in open waste areas. While *S. viridis* is an extensive seed producer, it is a poor competitor and only severely affects crop yields when the seedlings emerge

at about the same time as the crop (Blackshaw et al. 1981; Peterson and Nalewaja 1992). Over the past 20 years, green foxtail has acquired resistance to several groups of herbicides ([www.weedscience.org](http://www.weedscience.org)) and, as a result, has retained its status as one of top 10 most prevalent weeds in Western Canada.

*Setaria viridis* is often found in mixed populations with *S. faberi* Herrm. (Japanese bristlegrass) and *S. pumila* (Poir.) Roem. et Schult (yellow foxtail). The first record of *S. faberi*, giant foxtail, in the US was on Long Island in 1925 (Fairbrothers 1959). It is generally assumed that *S. faberi* was introduced as a seed contaminant from China where the species occurs in several regions as a common weed. It may have spread in the US along the railroads, and became a major agricultural problem in the Corn Belt in the 1950s (Knake 1990). Morphologically, *S. faberi* is very similar to and easily confused with *S. viridis* (Fairbrothers 1959; Layton and Kellogg 2014). The characteristics that best distinguish the two species are the greater length ratio of the lemma and 2nd glume in *S. faberi* compared to *S. viridis*, and the pubescence on the upper surface of the leaf blades of *S. faberi*. However, leaf pubescence is not universally present in *S. faberi* in China (Knake 1990). At a genetic level, *S. faberi* is an allotetraploid with one of the genomes being similar to that of the diploid *S. viridis*. Both species can cross to form triploids (Li et al. 1942; Willweber-Kishimoto 1962), and there is thus some potential for gene flow between the two species.

*S. pumila*, yellow foxtail, occurs in multiple ploidy forms ( $2n = 18, 36, 54$  and  $72$ ) (Rominger 1962; chromosome counts database—<http://ccdb.tau.ac.il>). The species is morphologically and genetically distinct from *S. viridis*. Phylogenetic analysis places *S. pumila* with African *Setaria* species, suggesting that *S. pumila* is native to Africa (Kellogg et al. 2009). Despite their different genome composition, a recent study reported obtaining seed from a cross between *S. viridis* and *S. pumila* (Jiang et al. 2013). Although Jiang and colleagues did not investigate the hybrid nature of the seed, and their finding contrasts with earlier reports of unsuccessful attempts to cross both species (Till-Bottraud et al. 1992; Willweber-Kishimoto 1962), we nevertheless need to consider the potential for gene flow between the two species.

Due to its small genome (510 Mb; <http://data.kew.org/cvalues/>), diploid nature ( $2n = 18$ ), and short life cycle, *S. viridis* and its domesticated form, *S. italica*,

have become important models to study the genetics of the biofuel crop switchgrass and for  $C_4$  photosynthesis (Brutnell et al. 2010; Doust et al. 2009; Li and Brutnell 2011). The genome of *S. italica* has been sequenced and assembled into 9 pseudomolecules, corresponding to the 9 chromosomes, and covers ~80 % of the genome and more than 95 % of the gene space (Benetzen et al. 2012; Zhang et al. 2012). The Joint Genome Institute (JGI) has also sequenced several accessions of *S. viridis* using the Illumina platform (sequencing reads available from NCBI's Sequence Read Archive (SRA)).

Several analyses of the diversity of *S. viridis* have been carried out. Wang et al. (1995) analyzed a set of 168 *S. viridis* accessions, some 75 % of which were collected in North America with the remaining 25 % distributed across Europe and Asia, with 13 isozyme markers. Their results indicated that *S. viridis* had been introduced multiple times from Eurasia into North America and formed two geographically separated subpopulations. Jia et al. (2013) analyzed a set of 288 mainly Chinese accessions with 77 SSR markers and also identified two subpopulations. Although the subpopulations did not correspond to geographical eco-regions, one of the subpopulations comprised mostly accessions from higher latitude eco-regions in Northern China while the majority of lines from lower latitude eco-regions formed the second subpopulation. In our study, we analyzed *S. viridis* and *S. faberi* accessions collected from the US and Canada and compared their DNA profiles obtained with 11 SSR markers with those of a global *Setaria* collection consisting mainly of *S. viridis* lines but also comprising some *S. italica* accessions.

## Materials and methods

### Plant materials

A total of 115 *S. viridis* accessions (232 lines), 11 *S. italica* accessions (11 lines), eight *S. faberi* accessions (22 lines), and one *S. verticillata* (L.) P. Beauv. (hooked bristlegrass) accession (1 line) were analyzed within this study. The term 'accession' is used in a broad sense and can represent a population of genetically dissimilar individuals growing at the same geographic location, as well as a sample of genetically identical genotypes. Forty-seven *S. viridis* and seven

*S. faberi* accessions were collected as part of this project in North America (US and Canada) and for each accession, one to five plants (referred to as 'lines') were sampled per location. A further 12 *S. viridis* accessions from Canada were obtained from Hugh Beckie, Agriculture and Agri-Food Canada, Saskatoon. In addition, 15 *S. viridis* accessions from the Middle East, seven from Western Europe, 28 from East Asia, three from South Asia, two from Central America and one from South America were obtained from various sources (Online Resource 1). A list of the accession numbers, species name, country of origin and where known, global positioning system (GPS) coordinates of the collection sites, and source of the seed are given in Online Resource 1.

For the accessions collected in situ as part of this project, single seeds from one to five sampled plants were planted, grown to maturity in the glasshouse under natural day light and selfed. Each of the sampled plants from a single location was given the same accession number with the suffix\_1, \_2, etc. For accessions obtained from other sources, one or two plants were grown for each accession. Again, these plants were given the same accession number with a number suffix.

### Differentiating *S. faberi* from *S. viridis*

*S. faberi* and *S. viridis* are morphologically highly similar and largely overlap in their vegetative characters. The main distinguishing characteristics are sparse pubescence on the adaxial side of the leaf blade and a short upper glume not longer than 90 % of the lower lemma length in *S. faberi* compared to glabrous leaves and a glume largely covering the seed in *S. viridis* (Fairbrothers 1959). All lines collected in situ by the authors had been identified as *S. viridis* or *S. faberi* prior to the DNA analysis. For the lines obtained from other sources, we used the ratio of the length of the upper glume to lower lemma as the criterion to identify *S. faberi* accessions that had been previously classified as *S. viridis*. Two accessions from China, 8125 and 81-79, had a glume length characteristic of *S. faberi*. Both accessions were considered as *S. faberi* in all analyses.

### Genotyping

DNA was extracted from approximately 10 mg of leaf tissue using a CTAB method (Doyle and Doyle 1987).

Each line was genotyped with 11 foxtail millet SSRs (p3, p16, p29, p88, p89, p95, b101, b102, b127, b163 and b166; Jia et al. 2009). PCR reactions were carried out in 15  $\mu$ L volumes comprising 50 ng of template DNA, 3  $\mu$ L 5 $\times$  buffer (Promega), 1.5 mM MgCl<sub>2</sub>, 200  $\mu$ M dNTPs, 67 nM M13-tailed specific forward primer, 267 nM fluorescently labeled M13 primer, 267 nM specific reverse primer and 0.9 U GoTaq<sup>®</sup> Flexi DNA Polymerase (Promega). The PCR conditions were 5 min denaturation at 95 °C, followed by 36 cycles of 15 s denaturation at 95 °C, 30 s annealing at 57 °C, and 30 s elongation at 72 °C. A final step of 10 min at 72 °C ensured full synthesis of PCR fragments. Three amplicons (3  $\mu$ L each) labeled with different fluorochromes were pooled and diluted three-fold. Three  $\mu$ L of the pooled and diluted amplicons were added to 7  $\mu$ L of a 19:1 mix of Hi-Di formamide (Applied Biosystems) and GeneScan 500 ROX-labeled marker (Applied Biosystems). Amplicons were size-separated on an ABI 3730xl (Applied Biosystems). SSR profiles were evaluated in GeneMarker<sup>®</sup> (Softgenetics) and peaks were scored manually.

## Data analysis

### Population structure analysis

A population structure analysis was conducted using the program STRUCTURE 2.3.1 (Pritchard et al. 2000). No prior information on the geographic origin of the lines was used. Lines that contained identical alleles for all SSRs tested were analyzed as a single entry, so that each entry in the population structure analysis had a different allele composition. Identification of duplicate allele patterns was done using GenAIEx 6.501 (Peakall and Smouse 2012). STRUCTURE was run with K-values varying from 1 to 20 in an admixture model with a burn-in phase of 100,000 iterations and 1,000,000 Markov Chain Monte Carlo (MCMC) iterations and 50 runs for each K. The most likely number of subpopulations (K) was estimated according to Evanno et al. (2005). For a given K, the run with the highest posterior probability (out of 50) was selected for analysis.

Additionally, we analyzed the data with the software InStruct (Gao et al. 2007) which, similarly to STRUCTURE, is a Bayesian clustering method but for inbreeding species. We tested K values from 2 to 20

with the same parameters as used for STRUCTURE, except that only seven iterations were done for each K.

### Principal coordinates and diversity analyses

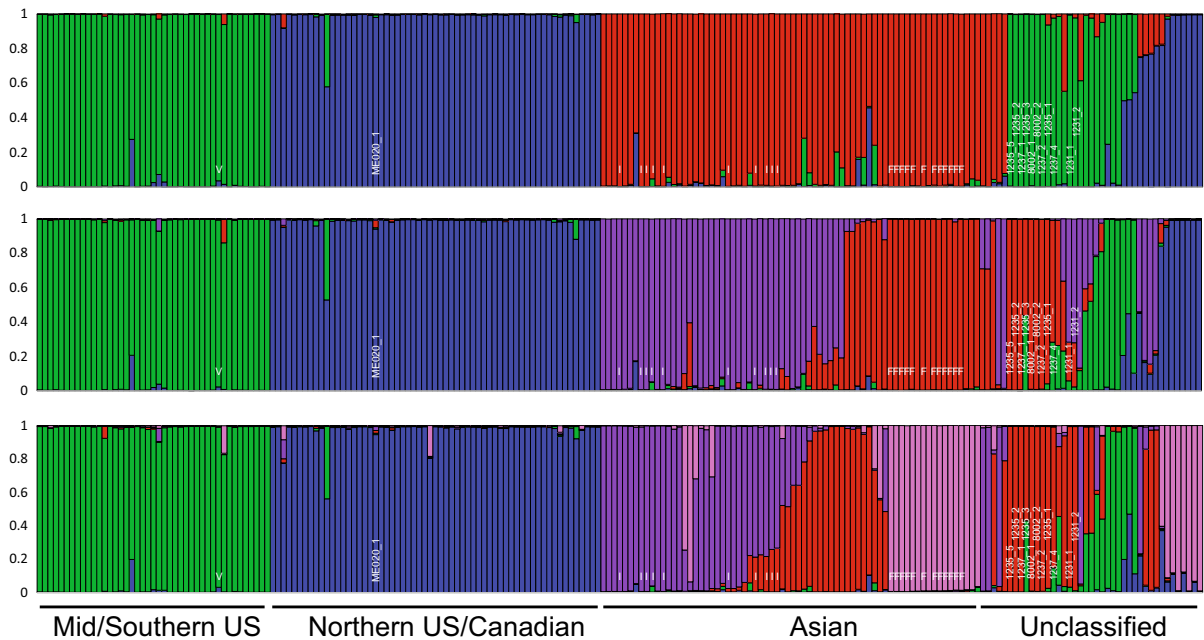
All analyses were carried out using GenAIEx version 6.501. The principal coordinates analysis (PCoA) was done with the Covariance—Standardized option (Peakall and Smouse 2012). Correlations between the genetic and the log(1 + geographic distance) transformed geographic distance of samples were analyzed using a Mantel test (Mantel 1967). Analyses of molecular variance (AMOVA) estimated the contribution of each locus to the total variance in each subpopulation and partitioned the total molecular variance within and between subpopulations.

### Neighbor-joining tree

The program MICROSAT v1.5 (Eric Minch, Stanford University, USA; <http://hpgl.stanford.edu/projects/microsat/>) was used to calculate genetic distances (Dps) based on the proportion of shared alleles (ps) with  $Dps = 1 - ps$ . The distance matrix was used as input in Phylip v 3.69 to generate a neighbor-joining tree (Felsenstein 1993; <http://evolution.genetics.washington.edu/phylip.html>). Sample inputs were randomized (J option) and the tree was rooted using *S. faberi* accession 81-79 as outgroup. For comparison, and to check the robustness of the clusters, 12 distance matrices based on shared alleles for (1) all 11 SSR markers and (2) subsets of 10 out of the 11 SSR markers and the corresponding trees were generated using PowerMarker V3.25 (Liu and Muse 2005).

### Summary statistics

GenAIEx version 6.501 (Peakall and Smouse 2012) was used to calculate the number of alleles ( $N_a$ ), effective alleles ( $N_e$ ) and specific alleles (alleles specific to a single subpopulation) in each subpopulation. In addition, the expected heterozygosity ( $H_e$ ) was calculated across the entire population and by subpopulation. Only accessions that (1) were a member of the same subpopulation with a membership  $\geq 95\%$  using both STRUCTURE and InStruct at  $K = 3$  and (2) retained their membership to the same overarching population group when K was increased from 3 to 6 were included in the calculations.



**Fig. 1** Structure plots for  $K = 3$ ,  $K = 4$  and  $K = 5$ . Genotypes represented in the plot are listed in the same order as in Online Resource 3. All genotypes are *S. viridis* except those indicated with ‘V’ (*S. verticillata*), ‘I’ (*S. italica*) and ‘F’ (*S. faberi*). Some

accessions discussed in the text are annotated by name. The ‘unclassified’ group consists of accessions that change subpopulation with varying  $K$ . (Color figure online)

## Results

### Population structure

#### *Population structure analysis*

A total of 266 lines were genotyped with 11 SSR markers. The majority of the lines (226 lines) had no missing data, 34 lines had missing data for 1 SSR and 6 lines had missing data for 2 SSRs. After grouping lines with identical genotypes at all 11 loci, a total of 192 *S. viridis* entries, 11 *S. faberi* entries, 11 *S. italica* entries and one *S. verticillata* entry were analyzed using STRUCTURE and InStruct. Lines with identical genotypes are given in Online Resource 2.

The  $\Delta K$  plot based on  $\text{LnP}(D)$  values from the STRUCTURE analysis indicated that the most likely number of populations ( $K$ ) was 3. InStruct, however, indicated that, based on the deviance information criteria (DIC), the optimal value of  $K$  was 17. Because the accuracy of various methods for determining the optimal number of subpopulations is greatly decreased when using a small number of markers (Gao et al. 2011), we conducted both STRUCTURE and InStruct

analyses with  $K$  varying from 3 to 6, and then manually assessed the composition of subpopulations as  $K$  increased (Fig. 1; Online Resource 3). For the subgroup descriptions below, we only considered accessions that belonged to the same subpopulations as determined by both STRUCTURE and InStruct, and that did not change membership across the three main groups identified at  $K = 3$  with increasing  $K$  value. At  $K = 3$ , the accessions largely separated into a Northern US/Canadian *S. viridis* group (latitudes above  $46^\circ\text{N}$ ), a Mid/Southern US *S. viridis* group which also comprised the *S. verticillata* accession (latitudes below  $44^\circ\text{N}$ ), and a predominantly Asian mixed group which contained *S. viridis* as well as the cultivated *S. italica* and the tetraploid wild species *S. faberi*. Subpopulation groupings largely corresponded with climatic zones (Online Resource 1). Increasing  $K$  to 4 placed all *S. faberi* accessions and a subset of the *S. viridis* accessions in a separate subgroup (Fig. 1). A further increase to  $K = 5$  reduced the number of *S. viridis* accessions associated with *S. faberi*. The *S. viridis* accessions that split off from *S. faberi*, together with some other *S. viridis* accessions, formed a separate subpopulation at  $K = 5$ . Increasing  $K$  to 6



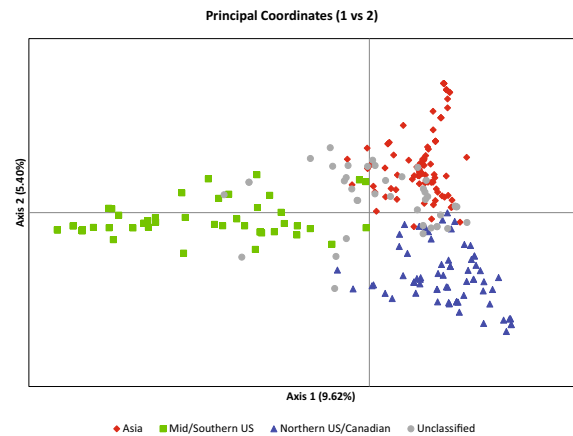
divided the Northern US/Canadian group into 2 subgroups (Online Resource 3). Overall, 176 of the 215 unique genotypes had membership to the same subpopulation at  $K = 3$  using both STRUCTURE and InStruct, and retained membership to those three overarching groups even when  $K$  was increased. The remaining 39 genotypes varied in their membership with varying  $K$  (Fig. 1; Online Resource 3). The change was largely unidirectional to the Asian group. For example, accessions 1231, 1235, 1237 and 8002 belonged to the Mid/Southern US *S. viridis* subpopulation at  $K = 3$  and to the Asian subpopulation at  $K = 4, 5$  and  $6$  (Fig. 1; Online Resource 3). Of the lines that switched population groupings as  $K$  was modified, 24 (62 %) were solidly assigned to a particular group at  $K = 3$  (membership value  $\geq 95$  %). Of these, 62 % belonged to the Mid/Southern US group, 29 % belonged to the Northern US/Canadian group, and 8 % belonged to the Asian group.

#### Principal coordinates analysis

The first three coordinates explained 19.64 % of the variation, with the first coordinate explaining 9.62 %, the second coordinate 5.40 % and the third coordinate 4.63 %. Color-coding of the accessions in the 2-dimensional PCoA plot showed a good correspondence between the population groups obtained from the PCoA and from the STRUCTURE/InStruct analyses (Fig. 2). Lines that changed membership with varying runs and/or  $K$  in the STRUCTURE/InStruct analyses largely grouped with the Asian accessions.

#### Neighbor-joining tree

Clustering of accessions in the neighbor joining tree generated with 11 SSR markers again was in good agreement with that obtained by the population structure and PCoA analyses (Fig. 3). To assess the robustness of the clustering, we generated distance matrices and trees based on subsets of 10 SSR markers and compared the tree topologies (Online Resource 4). Bar a few lines, the Mid/Southern accessions formed a single cluster in all trees. The Northern US/Canadian accessions largely grouped together in seven distance trees, and formed two clusters in four trees. The accessions belonging to the Asian subpopulation clustered to some extent, but these clusters were less



**Fig. 2** Principal coordinates analysis. Accessions are color coded according to the subpopulation to which they had majority (>50 %) membership based on STRUCTURE and InStruct analyses (*red diamonds* Asian subpopulation; *green squares* Mid/Southern US subpopulation; *blue triangles* Northern US/Canadian subpopulation; *gray circles* unclassified). (Color figure online)

robust than the North American/Canadian and Mid/Southern US clusters.

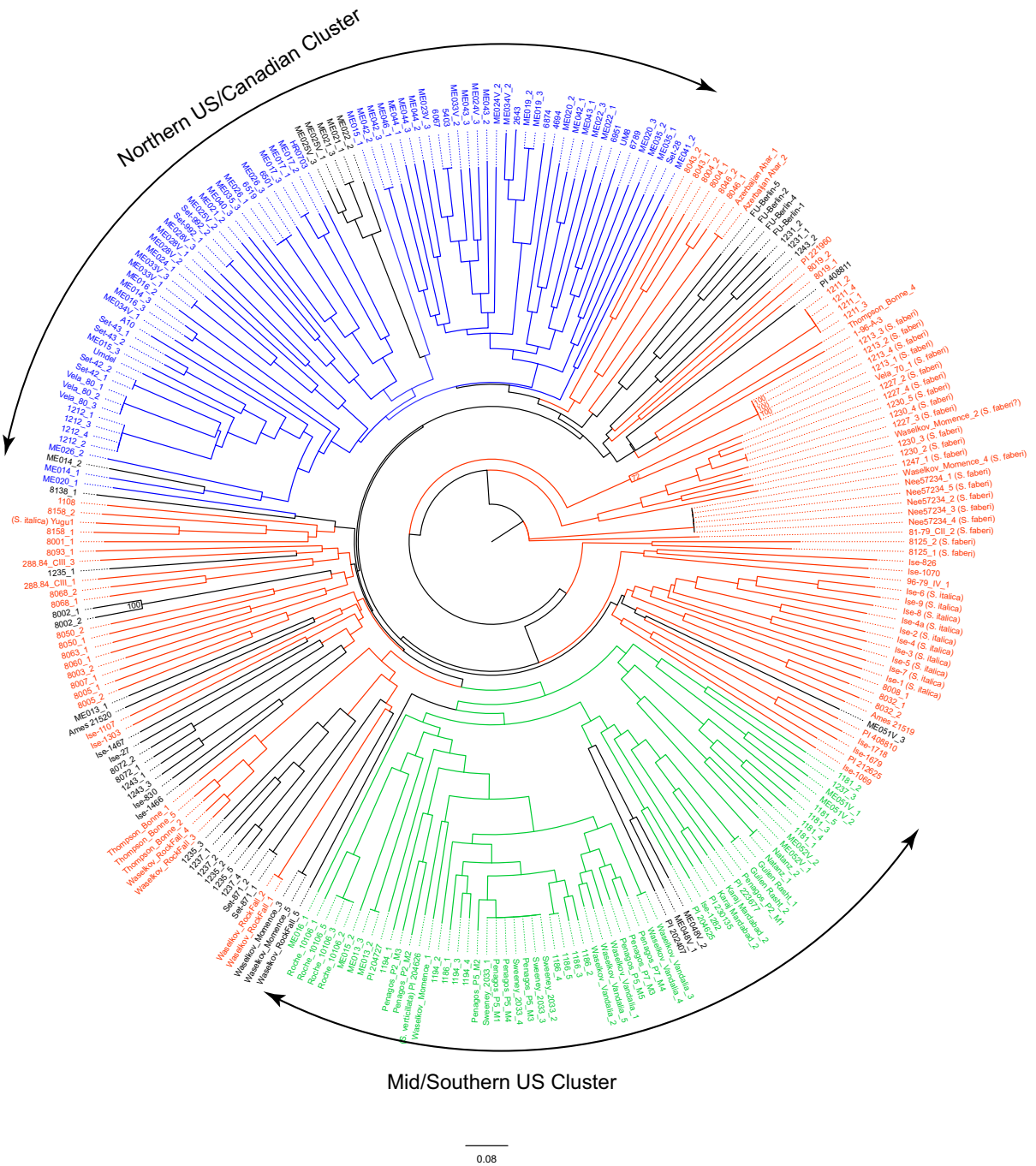
#### Overall genetic diversity

##### *Genetic diversity is correlated with geographic distance*

Significant positive correlations were found between the genetic distance and the geographic distance between accessions. This was true when all accessions were considered as well as for accessions within each of the three subpopulations. The correlation ( $R_{xy}$  value) was highest for accessions in the Asian subpopulation and lowest for accessions in the Northern US/Canadian subpopulation (Table 1).

##### *Genetic diversity by subpopulation*

All 11 SSR markers used showed polymorphisms in all three subpopulations, with the numbers of alleles varying from four (SSR p16 in the North American group) to 28 (SSR b101 in the Asian group) (Table 2). Both the mean number of alleles and effective number of alleles were lowest in the Mid/Southern US subpopulation (7.7; 2.5) and highest in the Asian subpopulation (20.4; 10.6) (Table 2). Interestingly, in the Mid/Southern US subpopulation, a single allele was present at a



**Fig. 3** Neighbor joining tree. Accessions are color coded according to the subpopulation to which they had majority (>50 %) membership based on STRUCTURE and InStruct

analyses (red Asian subpopulation; green Mid/Southern US subpopulation; blue Northern US/Canadian subpopulation; black unclassified). (Color figure online)

frequency >50 % for eight of the 11 loci analyzed (Table 2; Online Resource 5). For six of these loci, the allele that was predominant in the Mid/Southern

subpopulation was either absent or present at a frequency <5 % in both the Asian and Northern US/Canadian subpopulations. In the Northern US/Canadian

**Table 1** Correlation between genetic distance and geographic distance based on a Mantel test

Population	Rxy	P(rxy-rand $\geq$ rxy-data)
All	0.353	0.001
Asian	0.562	0.001
Mid/Southern US	0.338	0.001
Northern US/Canada	0.267	0.001

subpopulation, a predominant allele (frequency of 70.3 %) was found for a single SSR locus, p89, although this allele was also present at a frequency of 12.5 % in the Asian subpopulation (Online Resource 5). In contrast, the highest frequency of any single allele in the Asian subpopulation was 30.9 %. Perhaps not surprisingly, 50 % of all alleles in the Asian subpopulation were minor alleles (present in  $\leq 5$  % of the individuals within a given subpopulation) compared to 11 and 27 % of the alleles in the Mid/Southern US and Northern US/Canadian subpopulations, respectively.

On average, 71 % of the variation was found between individuals within subpopulations, and 29 % of the variation among subpopulations. The Asian subpopulation had the highest level of variation (expected heterozygosity ( $H_e$ ) of  $0.901 \pm 0.007$ ), followed by the Northern US/Canadian subpopulation

( $H_e = 0.755 \pm 0.036$ ) and the Mid/Southern US subpopulation ( $H_e = 0.512 \pm 0.064$ ).

## Discussion

Genome relationships between *S. viridis*, *S. faberi*, *S. verticillata* and *S. pumila*

While the focus of this study was on *S. viridis*, we also collected *S. faberi*, which is morphologically highly similar to and hence can easily be confused with *S. viridis*, and *S. pumila* which often grows in sympatry with *S. viridis*. Several *S. italica* accessions, the domesticated form of *S. viridis*, were also included in our analysis. In addition, a single *S. verticillata* accession was obtained from GRIN. The SSR markers, which were developed against *S. viridis* sequence data (Jia et al. 2009) amplified equally well using *S. viridis*, *S. italica*, *S. verticillata* and *S. faberi* genomic DNA as a template. However, most primer sets failed to amplify or produced weak and/or complex patterns in *S. pumila*. This is in agreement with the results of phylogenetic analyses that place *S. viridis*, *S. italica* and one genome of the tetraploids *S. faberi* and *S. verticillata* into a single clade while *S. pumila* is more distantly related (Kellogg et al. 2009; Layton and

**Table 2** Number of alleles (Na), effective alleles (Ne) and heterozygosity (He) overall and per subpopulation

Marker	Chromosomal location <sup>a</sup>	Asian				Mid/Southern				Northern US/Canada				All		
		Na	Ne	He	Spe	Na	Ne	He	Spe	Na	Ne	He	Spe	Na	Ne	He
p3	1 (40.75 Mb)	20	7.049	0.858	–	5	1.844	0.458	205	12	5.792	0.827	–	30	10.221	0.902
p29	7 (31.74 Mb)	24	13.501	0.926	–	7	2.854	0.650	180	10	5.432	0.816	–	37	13.518	0.927
b127	5 (28.40 Mb)	19	8.867	0.887	–	10	2.795	0.642	329	9	4.222	0.763	–	26	10.041	0.900
p88	1 (29.83 Mb)	21	11.306	0.912	–	7	1.781	0.438	217	9	4.261	0.765	–	29	11.368	0.912
b101	3 (4.07 Mb)	28	8.721	0.885	–	5	1.338	0.253	261	20	5.531	0.819	–	40	10.368	0.904
p95	9 (12.28 Mb)	21	12.117	0.917	–	6	1.532	0.347	–	11	4.859	0.794	–	24	6.475	0.846
p89	4 (30.40 Mb)	22	13.287	0.925	–	7	1.539	0.350	232	6	1.862	0.463	–	28	6.387	0.843
p16	1 (32.95 Mb)	15	8.591	0.884	–	6	1.267	0.211	–	4	2.530	0.605	–	17	4.976	0.801
b102	9 (5.90 Mb)	22	14.013	0.929	–	12	4.845	0.794	–	16	5.358	0.813	–	38	17.800	0.943
b166	9 (14.11 Mb)	19	11.620	0.914	–	12	4.658	0.785	–	16	7.027	0.858	–	27	13.679	0.927
b163	3 (47.64 Mb)	13	7.844	0.873	–	8	3.406	0.706	–	11	4.778	0.791	–	24	11.862	0.915
Mean		20	10.629	0.901		8	2.533	0.512		11	4.696	0.756		29	10.609	0.893

For each marker, the allele (in bp) that is present at a frequency of  $>50$  % in one of the subpopulations and  $<5$  % in both other subpopulations is indicated as specific alleles (Spe)

<sup>a</sup> The chromosomal location was obtained by conducting a Blastn search of the primer sequences listed in Jia et al. (2009) against the *S. italica* genome sequence (Bennetzen et al. 2012)



Kellogg 2014). Our analysis of 30 *S. pumila* accessions with 11 SSR markers yielded no evidence of gene flow between *S. pumila* and *S. viridis* growing in sympatry.

#### Relationships between accessions based on population structure and neighbor-joining tree

Eighty-nine percent of the *S. viridis* lines (93 % of the accessions) from regions in the US covered by our analysis (Online Resource 1) can be grouped into two subpopulations, referred to as a Mid/Southern US group and a Northern US/Canadian group (Figs. 1, 2). Groupings obtained using STRUCTURE/InStruct aligned completely with the relationships revealed by a neighbor joining analysis except for one line, ME020\_1, which was classified as belonging to the Northern US/Canadian subpopulation by STRUCTURE and InStruct, but grouped with Chinese accessions in the neighbor-joining tree (Fig. 3). ME020\_1 carried rare alleles at three loci and also had missing data at two loci which may explain its odd placement in the neighbor joining tree. ME020\_1 did, however, carry the 204 bp allele at locus p89, which is present in 69 % of the lines belonging to the Northern US/Canadian subpopulation. Interestingly, removal of SSR b127 from the analysis clustered ME020\_1 with the Northern US/Canadian subpopulation (Online Resource 4).

Accessions that belong to the Mid/Southern US subpopulation are found mainly below latitudes 44°N in Köppen's climate zone Cfa (warm temperate, fully humid, hot summer) (Kottek et al. 2006), while accessions belonging to the Northern US/Canadian subpopulation are found mainly above latitude 46°N in climate zone Dfb (snow climate, fully humid, warm summer). These results are in agreement with observations made by Wang et al. (1995) based on isozyme analysis of US *S. viridis* accessions and by Huang et al. (2014) based on close to 40,000 single nucleotide polymorphisms. The majority of the accessions from Iran and Turkey (climate zones Csa—warm temperate, dry hot summer; Csb—warm temperate, dry warm summer) grouped with the Mid/Southern US subpopulation suggesting that Mid/Southern US accessions may be derived from introductions from Southern Europe and/or the Middle East. The number of introductions with favorable allele combinations that gave rise to the Mid/Southern population may have

been limited as indicated by the presence of predominant alleles at eight of the 11 SSR loci. The single *S. verticillata* line that was included in our analysis originated from Turkey and also grouped with the Mid/Southern US lines in both the population structure and neighbor joining analyses. The Western European accessions we analyzed (climate zone Cfb—warm temperate, fully humid, warm summer) largely grouped with the Northern US/Canadian subpopulation, suggesting that introductions from Western Europe may have given rise to the Northern US and Canadian *S. viridis* populations. Because a predominant allele is found only at a single SSR locus, the Northern US/Canadian subpopulation probably originated from a larger number of introductions than the Mid/Southern subpopulation. The observed groupings likely reflect the differential adaptation of Turkish, Iranian and Mid/Southern US lines to dry or humid hot summers with <15 h day lengths on one hand, and the Western European and Northern US/Canadian lines to humid and warm (but not hot) summers with >15 h day lengths on the other hand. The SSR with the predominant allele in the Northern US/Canadian subpopulation is located on foxtail millet chromosome IV in a region that carries a flowering time QTL and may be associated with adaptation to Northern climates. The Mid/Southern US subpopulation has a different predominant allele at this locus. Differential adaptation to environmental conditions at different latitudes could be observed clearly when Canadian accessions were grown in the glasshouse in Georgia. Most accessions flowered very early, yielding mature plants that were very small and set little seed. This plant phenotype was very different from that observed when the plants grew in their native environment.

In contrast to the two largely North American subpopulations that formed distinct clusters in the neighbor-joining tree, lines belonging to the Asian subpopulation largely fell into three clusters (Fig. 3), but membership to the clusters varied depending on the software used to generate the distance matrices and trees, and on the subset of SSRs used in the analysis (Online Resource 4). This can likely be explained by the diversity of the accessions that formed the Asian subpopulation. No alleles were identified in the Asian subpopulation that were present in 50 % or more of the accessions, and 50 % of the alleles were minor alleles. This is in contrast to the Mid/Southern US population which had low genetic variation with predominant

alleles at 72 % of the SSR loci leading to stable clustering of accessions across different analyses.

In addition to *S. viridis*, our study included eight *S. faberi* accessions and 11 *S. italica* accessions, all of which belonged to the Asian subpopulation. The 10 *S. italica* accessions from India consistently clustered together and were more closely related to *S. viridis* accessions from India and Afghanistan than to Yugu1, a *S. italica* accession from China (Fig. 3). The co-grouping of *S. italica* with *S. viridis* from the same geographic region is consistent with previously published data (Le Thierry d'Ennequin et al. 2000) and indicates that gene flow exists between the two species. The *S. faberi* accessions also consistently clustered together. It had previously been suggested that *S. faberi* was introduced into the US from China (Rominger 1962). While membership of the *S. faberi* lines to the Asian subpopulation seems to support this, the fact that *S. faberi* accessions formed a separate cluster in the neighbor joining tree and that increasing K from 3 to 4 in the STRUCTURE and InStruct analyses resulted in the splitting off of the *S. faberi* accessions and some *S. viridis* lines from the Asian subpopulation indicates that this interpretation needs to be treated with caution. The *S. viridis* lines that grouped with *S. faberi* at K = 4 belonged to four accessions from the US, four accessions from China, one accession from Germany and one accession from Iran, which is too small a dataset to determine a country bias. In our collection of *Setaria* accessions, we had one accession (Waselkov\_Momence) for which, of the five collected lines, one was classified as *S. faberi* based on glume size while the others were confirmed as *S. viridis*. Interestingly, one of the *S. viridis* lines (Waselkov\_Momence 2) also fell into the *S. faberi* cluster in the neighbor-joining tree. Waselkov\_Momence\_2 carried alleles that were common in *S. faberi* at nine of the 11 loci. However, at two of the loci, the alleles in Waselkov\_Momence 2 were absent from any of the *S. faberi* lines but were present in *S. viridis*, one at a low frequency in the Asian (~8 %) and Northern US/Canadian subpopulation (~3 %), and the other at high (~80 %) and moderate (~20 %) frequencies in the Mid/Southern US subpopulation and Northern US/Canadian subpopulation, respectively. *S. faberi* is an allotetraploid with one genome donated by *S. viridis*, so it is not surprising that the two species should share alleles (Benabdelmouna et al. 2001; Layton and Kellogg 2014).

Although the diversity analysis was conducted with 11 SSRs only, the results of the population structure analyses were highly similar to those obtained by Huang et al. (2014) who analyzed a largely overlapping set of germplasm using ~40,000 SNP markers obtained using genotyping-by-sequencing (GBS). Of the 112 lines that were in common between the two studies and that consistently grouped within the same subpopulation at K = 3 in the SSR study, all but two (98 %) were classified in the same subpopulation using SNPs and SSRs. For the purpose of comparison, lines were classified to the subpopulation in which they had >50 % membership. The GBS data, however, indicated a higher percentage of admixed lines (<90 % membership to a single subpopulation) than the SSR analysis (33 vs. 6 %). Lines that were classified as admixed in the SSR analysis were also classified as admixed in the SNP analysis. The only exceptions were the lines Azerbaiyan Ahar and PI 221960 which were classified as admixed with a majority membership to the Northern US/Canadian subpopulation in the GBS study but belonged to the Asian subpopulation in the SSR analysis. These were the only two accessions in our study that originated from climate zone Dsb (snow climate with dry warm summers). When considering the 24 lines that were in common between the SNP and SSR studies, but did not have a consistent membership to a particular subpopulation using different software packages and/or different K values based on the 11 SSRs, 54 and 79 % of classifications agreed between the GBS data and the SSR results obtained at K = 3 and K = 4, respectively. At K = 4, the Asian group largely split into two subgroups, and several of the lines that were classified as Mid/Southern US in the SSR analysis at K = 3 belonged to one of the two Asian subgroups at K = 4. The lines that differed in their classification between GBS SNP and SSR data were typically highly admixed (membership to a single subpopulation was <60 %).

## Conclusions

Small numbers of SSR markers can be sufficient, depending on the species and genetic diversity present, to determine the overall population structure of germplasm collections. To solidly determine population groups, especially at low marker numbers, it can

be helpful to assess the stability of line classifications at increasing K values. Lines that change membership may be more likely to have high levels of undetected admixture. Our analysis demonstrated that *S. viridis* lines in the US were likely introduced from Europe and/or the Middle East. The fact that Northern US and Canadian populations have a closer genetic relationship to *S. viridis* populations from Western Europe, and Mid/Southern US populations have a closer genetic relationship to *S. viridis* populations from Southern Europe and the Middle East suggests that *S. viridis* will only flourish if introduced to the climatic and/or photoperiod zones from which it originates and to which it is adapted.

**Acknowledgments** We thank D. Vela, K. Waselkov, J. Thompson, P. Sweeney, C. Roché, J. Penagos, M. Weigend, H. Beckie, T. Robert, M. Keshavarzi, A. Börner, USDA and ICRISAT for sharing of *Setaria* seed. This work was funded by National Science Foundation awards DEB-0952177 and DEB-0952185 to KMD and EAK, respectively.

#### Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

#### References

- Benabdelmouna A, Shi Y, Abirached-Darmency M, Darmency H (2001) Genomic in situ hybridization (GISH) discriminates between the A and the B genomes in diploid and tetraploid *Setaria* species. *Genome* 44:685–690
- Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn J, Grimwood J, Jenkins J, Barry K, Lindquist E, Hellsten U, Deshpande S, Wang X, Wu X, Mitros T, Triplett J, Yang X, Ye C-Y, Mauro-Herrera M, Wang L, Li P, Sharma M, Sharma R, Ronald PC, Panaud O, Kellogg EA, Brutnell T, Doust AN, Tuskan GA, Rokhsar D, Devos KM (2012) Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol* 30:555–561
- Blackshaw RE, Stobbe EH, Sturko ARW (1981) Effect of seeding dates and densities of green foxtail (*Setaria viridis*) on the growth and productivity of spring wheat (*Triticum aestivum*). *Weed Sci* 29:212–217
- Brutnell TP, Wang L, Swartwood K, Goldschmidt A, Jackson D, Zhu XG, Kellogg E, Van Eck J (2010) *Setaria viridis*: a model for C-4 photosynthesis. *Plant Cell* 22:2537–2544
- Darmency H (2005) Incestuous relations of foxtail millet (*Setaria italica*) with its parents and cousins. *Crop fertility and volunteerism*. CRC Press, Boca Raton, pp 81–96
- Douglas BJ, Gordon TA, Morrison IN, Maw MG (1985) The biology of Canadian weeds. 70. *Setaria viridis* (L.) Beauv. *Can J Plant Sci* 65:669–690
- Doust AN, Kellogg EA, Devos KM, Bennetzen JL (2009) Foxtail millet: a sequence-driven grass model system. *Plant Physiol* 149:137–141
- Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull* 19:11–15
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620
- Fairbrothers DE (1959) Morphological variation of *Setaria faberii* and *S. viridis*. *Brittonia* 11:44–48
- Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle
- Forcella F, Harvey SJ (1983) Relative abundance in an alien weed flora. *Oecologia* 59:292–295
- Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* 176:1635–1651
- Gao H, Bryc K, Bustamante CD (2011) On identifying the optimal number of population clusters via the deviance information criterion. *PLoS ONE* 6:e21014
- Groh H, Frankton C (1949) Canadian weed survey. 7th report. Canada Department of Agriculture, Ottawa, p 144
- Huang P, Feldman M, Schroder S, Bahri BA, Diao X, Zhi H, Estep M, Baxter I, Devos KM, Kellogg EA (2014) Population genetics of *Setaria viridis*, a new model system. *Mol Ecol* 20:4912–4925
- Jia X, Zhang Z, Liu Y, Zhang C, Shi Y, Song Y, Wang T, Li Y (2009) Development and genetic mapping of SSR markers in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *Theor Appl Genet* 118:821–829
- Jia G, Shi S, Wang C, Niu Z, Chai Y, Zhi H, Diao X (2013) Molecular diversity and population structure of Chinese green foxtail [*Setaria viridis* (L.) Beauv.] revealed by microsatellite analysis. *J Exp Bot* 64:3645–3656
- Jiang H, Barbier H, Brutnell T (2013) Methods for performing crosses in *Setaria viridis*, a new model system for the grasses. *J Vis Exp* 80:e50527
- Kellogg EA, Aliscioni SS, Morrone O, Pensiero J, Zuloaga F (2009) A phylogeny of *Setaria* (Poaceae, Panicoideae, Paniceae) and related genera, based on the chloroplast gene *ndhF*. *Int J Plant Sci* 170:117–131
- Knake EL (1990) Giant foxtail: *Setaria faberi* Herrm. Bulletin of the University of Illinois at Urbana-Champaign, College of Agriculture, Agricultural Experiment Station 803
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. *Meteorol Z* 15:259–263
- Layton DJ, Kellogg EA (2014) Morphological, phylogenetic, and ecological diversity of the new model species *Setaria*

- viridis* (Poaceae: Paniceae) and its close relatives. *Am J Bot* 101:539–557
- Le Thierry d'Ennequin M, Panaud O, Toupance B, Sarr A (2000) Assessment of genetic relationships between *Setaria italica* and its wild relative *S. viridis* using AFLP markers. *Theor Appl Genet* 100:1061–1066
- Li P, Brutnell T (2011) *Setaria viridis* and *Setaria italica*, model genetic systems for the Panicoid grasses. *J Exp Bot* 62:3031–3037
- Li CH, Pao WK, Li HW (1942) Interspecific crosses in *Setaria*. II. Cytological studies of interspecific hybrids involving: 1. *S. faberii* and 2. a three way cross, F<sub>2</sub> of *S. italica* × *S. viridis* and *S. faberii*. *J Hered* 33:351–355
- Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21:2128–2129
- Manson JM (1932) Weed survey of the Prairie Provinces. Dominion of Canada, p 34
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Res* 27:209–220
- Naciri Y, Darmency H, Belliard J, Dessaint F, Pernès J (1992) Breeding strategy in foxtail millet, *Setaria italica* (L.P. Beauv.), following interspecific hybridization. *Euphytica* 60:97–103
- Peakall R, Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Peterson DE, Nalewaja JD (1992) Green foxtail (*Setaria viridis*) competition with spring wheat (*Triticum aestivum*). *Weed Technol* 6:291–296
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Qie L, Jia G, Zhang W, Schnable J, Shang Z, Li W, Liu B, Li M, Chai Y, Zhi H, Diao X (2014) Mapping of quantitative trait locus (QTLs) that contribute to germination and early seedling drought tolerance in the interspecific cross *Setaria italica* × *Setaria viridis*. *PLoS ONE* 9:e101868
- Rominger JM (1962) Taxonomy of *Setaria* (Graminaceae) in North America. University of Illinois Press, Urbana
- Till-Bottraud I, Reboud X, Brabant P, Lefranc M, Rherissi B, Vedel F, Darmency H (1992) Outcrossing and hybridization in wild and cultivated foxtail millets: consequences for the release of transgenic crops. *Theor Appl Genet* 83:940–946
- Wang RL, Wendel JF, Dekker JH (1995) Weedy adaptation in *Setaria* spp. I. Isozyme analysis of genetic diversity and population genetic structure in *Setaria viridis*. *Am J Bot* 82:308–317
- Wang T, Fleury A, Ma J, Darmency H (1996) Genetic control of dinitroaniline resistance in foxtail millet (*Setaria italica*). *J Hered* 87:423–426
- Willweber-Kishimoto E (1962) Interspecific relationships in the genus *Setaria*. *Control Biol Kyoto Univ* 14:1–41
- Zhang G, Liu X, Quan Z, Cheng S, Xu X, Pan S, Xie M, Zeng P, Yue Z, Wang W, Tao Y, Bian C, Han C, Xia Q, Peng X, Cao R, Yang X, Zhan D, Hu J, Zhang Y, Li H, Li N, Wang J, Wang C, Wang R, Guo T, Cai Y, Liu C, Xiang H, Shi Q, Huang P, Chen Q, Li Y, Wang J, Zhao Z, Wang J (2012) Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat Biotechnol* 30:549–554