Title

High-Density Single Nucleotide Polymorphism Linkage Maps of Lowland Switchgrass using Genotyping-by-Sequencing

Journal

The Plant Genome, 8(2)

ISSN

1940-3372

Authors

Fiedler, Jason D

Lanzatella, Christina

Okada, Miki

et al.

Publication Date

2015-07-01

DOI

10.3835/plantgenome2014.10.0065

Peer reviewed

Published July 10, 2015

ORIGINAL RESEARCH

High-Density Single Nucleotide Polymorphism Linkage Maps of Lowland Switchgrass using Genotyping-by-Sequencing

Jason D. Fiedler, Christina Lanzatella, Miki Okada, Jerry Jenkins, Jeremy Schmutz, and Christian M. Tobias*

Abstract

Switchgrass (*Panicum virgatum* L.) is a warm-season perennial grass with promising potential as a bioenergy crop in the United States. However, the lack of genomic resources has slowed the development of plant lines with optimal characteristics for sustainable feedstock production. We generated high-density single nucleotide polymorphism (SNP) linkage maps using a reduced-representation sequencing approach by genotyping 231 F_1 progeny of a cross between two parents of lowland ecotype from the cultivars Kanlow and Alamo. Over 350 million reads were generated and aligned, which enabled identification and ordering of 4611 high-quality SNPs. The total lengths of the resulting framework maps were 1770 cM for the Kanlow parent and 2059 cM for the Alamo parent. These maps show collinearity with maps generated with polymerase chain reaction (PCR)-based simple-sequence repeat (SSR) markers, and new SNP markers were identified in previously unpopulated regions of the genome. Transmission segregation distortion affected all linkage groups (LGs) to differing degrees, and ordering of distorted markers highlighted several regions of unequal inheritance. Framework maps were adversely affected by the addition of distorted markers with varying severity, but distorted maps were of higher marker density and provided additional information for analysis. Alignment of these linkage maps with a draft version of the switchgrass genome assembly demonstrated high levels of collinearity and provides greater confidence in the validity of both resources. This methodology has proven to be a rapid and cost-effective way to generate high-quality linkage maps of an outcrossing species.

Published in The Plant Genome 8
doi: 10.3835/plantgenome2014.10.0065

5585 Guilford Rd., Madison, WI 53711 USA
An open-access publication

SINGLE NUCLEOTIDE POLYMORPHISMS are common mutations that occur between related genomes and are increasingly becoming used as markers for basic research and genetic mapping (Vignal et al., 2002; Yang et al., 2010; Mammadov et al., 2012). The technology for scoring SNPs has advanced from low-throughput direct sequencing and high-resolution melting (Chagné et al., 2008) to microarrays that simultaneously score up to hundreds of thousands of markers by virtue of hybridization of a displayed DNA fragment (LaFramboise, 2009; Ganal et al., 2012). Various plant species have been genotyped and mapped with microarray methods (Ganal et al., 2012; Colasuonno et al., 2013), but a predetermined set of polymorphisms is required. Since such SNPs are typically species specific (Micheletti et al., 2011), expensive arrays need to be developed for every organism in question. This is not economically feasible for minor crops and nonmodel organisms or realistic for the study of new species with unknown SNP information. With the advent of advanced DNA sequencing technology, another generation of protocols to identify SNPs has emerged that leverages the speed and ease of

J.D. Fiedler, C. Lanzatella, and C.M. Tobias, USDA–ARS, Western Regional Research Center, 800 Buchanan St., Albany, CA 94710; M. Okada, Univ. of California–Davis, 1 Shields Ave., Davis, CA 95616; J. Jenkins and J. Schmutz, HudsonAlpha Genome Sequencing Center, 601 Genome Way, Huntsville, AL 358206, and Dep. of Energy, Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek, CA 94598. Received 9 Oct. 2014. Accepted 5 Mar. 2015. *Corresponding author (christian.tobias@ars.usda.gov).

Abbreviations: BWA, Burrows–Wheeler Aligner; GBS, genotype-by-sequencing; GQ, genotype quality; InDel, insertion–deletion; LCS, longest common subsequence; LG, linkage group; LOD, logarithm of odds; ML, maximum-likelihood; PCR, polymerase chain reaction; QC, quality control; SDA, single-dose alleles; SNP, single nucleotide polymorphism; SSR, simple-sequence repeat; TRD, transmission ratio distortion.

generating a large number of short DNA reads (Nielsen et al., 2011; Davey et al., 2011; Kopecký and Studer, 2013). These novel protocols include steps to direct the fixed sequencing capacity to specific sites of the genome, typically by initiating the sequencing reads from restriction endonuclease recognition sites (Davey et al., 2011). This reduction of the genome complexity allows sequencing of many individuals simultaneously via multiplexed barcoding (Baird et al., 2008) and achievement of sufficient depth of coverage to identify high-quality SNP markers (Glaubitz et al., 2014). These methods have been useful for plant genotyping (Hyten et al., 2010; Mammadov et al., 2012; Romay et al., 2013) and for the production of crop linkage maps (Poland et al., 2012; Spindel et al., 2013). The progression and differences of the various genotype-by-sequencing (GBS) protocols have been previously reviewed (Nielsen et al., 2011; Davey et al., 2011), with the method described by Elshire and colleagues being widely adopted (Elshire et al., 2011).

Here we apply a GBS sequencing approach to study the warm-season perennial grass, switchgrass, which shows promise as a dedicated bioenergy crop (McLaughlin and Adams Kszos, 2005; Yuan et al., 2008; Wright and Turhollow, 2010). An integral part of the native North American tall-grass prairie, drought-resistant switchgrass has traditionally been grown for forage, erosion control, and restoration (Parrish and Fike, 2005). Tetraploid and octoploid cytotypes are prevalent throughout North America (Triplett et al., 2012), with little intercrossing detected between the different ploidy levels (Hultquist et al., 1997; Narasimhamoorthy et al., 2008). The ecotype (upland or lowland) generally describes local geographic adaptation and most of the current bioenergy feedstocks being developed are lowland tetraploid ($2n = 4x = 36$) due to their higher yield potential (Parrish and Fike, 2005; Monti et al., 2008; Bhandari et al., 2010). Switchgrass is wind pollinated and possesses a prezygotic self-incompatibility system (Martínez-Reyna and Vogel, 2002) resulting in high levels of outcrossing. However, self-pollination has been observed in a small number of individuals (Liu and Wu, 2012; Liu et al., 2012). The subgenomes are sufficiently divergent to allow identification of unique markers on every chromosome, and inheritance is disomic in tetraploids (Missaoui et al., 2005; Okada et al., 2010). Genetic linkage maps based on restriction fragment length polymorphisms and SSR markers have been generated with two different switchgrass populations and progeny from a self-pollinated individual (K5 × A4, AP13 × VS16, and NL94) (Missaoui et al., 2005; Okada et al., 2010; Liu et al., 2012; Serba et al., 2013). These populations were integrated to create a four-founder mapping population to further elucidate switchgrass genetic structure (Li et al., 2014). An early draft genome assembly (v1.1) has been released by the US Department of Energy's Joint Genome Institute, with approximately half of the assembled DNA contigs scaffolded into pseudomolecules based on linkage maps from an AP13 × VS16 (lowland × upland) population as part of the effort to produce a high quality switchgrass genomic resource (Casler et al., 2011). However, genetic resources for trait analysis in switchgrass need to be further developed to facilitate basic and applied research goals.

These goals include advances in production in marginal areas, reduced N requirements, improved yield (Carroll and Somerville, 2009), enhanced cell-wall digestibility (Schmer et al., 2008), and winter hardiness (Casler et al., 2007). Marker-assisted selection (Collard and Mackill, 2008) can accelerate the breeding process to achieve these goals but only after genetic dissection of important traits occurs. High-density linkage maps are useful tools to directly identify sets of closely linked markers surrounding loci contributing to phenotypic variation with breeding potential.

We report here the construction of high-density linkage maps of two lowland tetraploid germplasms using GBS. Simple-sequence repeat and sequence-tagged polymorphism type markers are also available for this mapping population (Tobias et al., 2006) and were included in this study. This integration allows direct comparison of maps generated from the two different technologies (Okada et al., 2010). Direct comparison of these maps with the early-release draft genome allows indirect comparison to other linkage maps used to anchor the assembly and provides additional confidence of marker ordering. A large number of transmission segregation distorted markers were added to the map, identifying several regions of potential segregation distortion and multilocus interactions. A detailed analysis of comparisons between these different linkage maps highlights effects that distorted markers have on linkage maps to increase genetic information but also add uncertainty to marker placement.

Materials and Methods

Mapping Population

A full-sib mapping population was created from a cross between two lowland tetraploid ($2n = 4x = 36$)individuals (K5 x A4) and is fully described elsewhere (Okada et al., 2010). Initial genotyping identified approximately 5% of the population as self-pollinated or dihaploid and these were removed from further mapping analysis. The remaining 231 F_1 plants were used for mapping in this study.

DNA Extraction and Genotyping-by-Sequencing Library Construction

Genotyping-by-sequencing library production followed a previously established protocol (Elshire et al., 2011) with a few modifications. Briefly, parental and F_1 progeny total genomic DNA was extracted from young leaves dried in silica using a cetyltrimethylammonium bromide (CTAB) isolation procedure (Chen and Ronald, 1999). DNA was digested with *PstI* and ligated to Illumina adapters with a custom set of sequence index tags. Barcoded DNA was subsequently adapted with Illumina sequencing primers and purified with a Qiagen PCR cleanup kit. DNA libraries were pooled (96 samples per flow-cell lane) and

single-end reads were generated on an Illumina HiSeq 2000 (JCVI). The entire mapping population and both parents were sequenced on three lanes of a flow cell along with other switchgrass samples.

Illumina Read Analysis

Single reads were de-multiplexed, the barcodes removed, and trimmed to 92 bases with the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Initial quality control (QC) analysis was performed on all individuals with FASTQC. In the initial mapping cross, two spontaneous dihaploid individuals were identified (Young et al., 2010). One of these individuals (ALB280) was sequenced in the same manner as the rest of the population. Reads from this individual were clustered with USEARCH (Edgar, 2010) to generate sequence reference tags. After filtering out sequences that aligned to the chloroplast genome, 179,275 92-bp reference tags (16.5 MB) were kept. These sequences were used as the reference for the alignment of parent and F_1 progeny GBS sequences with Burrows-Wheeler Aligner (BWA) (option-samse) (Li and Durbin, 2010). Single nucleotide polymorphisms and insertion–deletion (InDel) markers were called with samtools and bcftools (v.0.1.19) using default options and Bayesian interference (bcftools option-c) (Li et al., 2009). Polymorphisms were filtered using vcftools (Danecek et al., 2011).

Segregation and Linkage Analysis

As an outcrossing plant, switchgrass exhibits a large degree of allelic heterozygosity. Thus, the linkage analysis was conducted using polymorphic single-dose alleles (SDAs) (Wu et al., 1992). Single-dose alleles were identified on the basis of a heterozygous genotype in one parent and homozygous genotype in the other parent and a goodness of fit to a 1:1 segregation ratio in the F_1 population using the χ^2 test ($\alpha = 0.05$). Polymorphisms that possessed SDA parent alleles but showed transmission ratio distortion (TRD) in the F_1 population were labeled as TRD-SDAs. We further differentiated the TRD-SDAs by the severity of distortion as measured with the χ^2 test: mild TRD-SDA for $0.001 \leq p < 0.05$ and severe TRD-SDA for $1 \times 10^{-15} \leq p < 0.001$.

The GBS data set was merged with the SSR genotype data set used to create prior linkage maps (Okada et al., 2010) and was analyzed with JoinMap4 (van Ooijen, 2006) using the outbreeder full-sib family as the population type. Maternal and paternal maps were analyzed separately following the two-way pseudo-testcross strategy (Grattapaglia and Sederoff, 1994). This analysis allows for the production of linkage maps without the requirement of backcrossed F_1 individuals by identifying SDAs inherited from the mother and father separately. The SDAs were grouped into LGs at the minimum independence test logarithm of odds (LOD) score of 10.0. The TRD-SDAs and SDAs missing more than 23 genotypes (10%) were excluded during construction of the framework maps. Loci within LGs were ordered using the maximum-likelihood (ML) mapping algorithm (Jansen et al., 2001). Separate SDAs within a LG with equivalent or nearly equivalent map positions with regard to the variation in the data (Jansen et al., 2001) were binned together and map order was then determined using only one SDA from each bin. The grouped SDAs that were not used in the linkage analysis were treated as accessory loci to those that were mapped. The SDAs were placed on the map using the regression mapping algorithm (Stam, 1993), with the relative order obtained above fixed at the minimum LOD score of 8.0 and maximum recombination fraction of 0.35. The Kosambi mapping function was used to obtain map distance (Kosambi, 1943).

Two different distorted maps were created. Mild TRD maps include the SDAs, mild TRD-SDAs, and SDAs with 10 to 20% missing data. Severe TRD maps include all the available markers with SDA parent genotypes and less than 20% missing data. The TRD maps were created by adding in the additional loci to the framework maps with a strongest cross-link parameter threshold of 10.0. The use of the strongest cross-link parameter to include the TRD-SDAs was necessary to maintain the integrity of each LG. If TRD-SDAs were included in the initial grouping phase, JoinMap4 returned a small number of poorly linked LGs that did not accurately represent the number of chromosomes. These maps were ordered and map distances were estimated as above. In many of the distorted maps, loci placement was poor, as determined by the probability and stress values returned by JoinMap4. Thus, all distorted maps were reordered using the regression-mapping algorithm without use of a ML starting order, and maps that contained the most likely loci placement were used. Additionally, markers with an average probability of correct placement less than 10% were removed from the maps, and LGs were reordered to remove any distorted-marker influence on adjacent markers. Approximately 2% of markers were removed from mild TRD maps, and ~7% were removed from severe TRD maps. These markers were designated as accessory markers to the map. Pairwise linkage of all the markers was determined by grouping the entire data set at a LOD of 3.0.

Linkage groups were named by identification of shared markers between GBS maps and prior SSR linkage maps which correspond to syntenic diploid chromosomes of foxtail millet [*Setaria italica* (L.) P. Beauv. subsp. *italica*] with arbitrary "a" and "b" subgenome designations. Linkage maps were plotted against the physical map by aligning reference tags that contained GBS or expressed sequence tag (EST)-SSR markers to the early-release draft genome assembly of *P. virgatum* (v1.1, Phytozome.net) with BLASTn. The majority (85%) of the loci on a LG that aligned to the v1.1 assembly did so to the same pseudomolecule with an *E*-value $<1 \times 10^{-30}$. If the best alignment did not correspond to the most likely pseudomolecule, it typically aligned to unassembled contigs or to the homeologous pseudomolecule. In these cases,

next-best alignments that corresponded to the likely pseudomolecule were used (E-value $<1 \times 10^{-5}$). Regions of segregation distortion on the mild TRD maps were designated by presence of at least three contiguous TRD-SDAs. Regions of distortion on the severe TRD maps were visualized with an eight-marker sliding-scale average of the χ^2 value plotted vs. the LG position. This sliding-scale value was capped to a value corresponding to severe distortion ($p < 0.001$). Regions of segregation distortion were determined by interrogation of this sliding-scale value in addition to the χ^2 values; highlighted regions represent at least three contiguous TRD-SDAs. Map comparisons were drawn with MapChart (Voorrips, 2002). Nonparametric test of independence was performed with two-sample Wilcoxon Rank Sum Test (Mann–Whitney test) wilcox.test command in R (Bauer, 1972).

Comparison

Individual maps were compared based on their shared marker interval distance and for marker order based on their longest common subsequence (LCS) using the qualV package (Jachner et al., 2007) in R. Results were expressed as markers present in the LCS as a fraction of the number of markers in common between maps.

Results

Short-Read Analysis

Illumina sequencing generated a large number of high-quality multiplexed single-end reads. At least 1 million 92-bp reads were generated for the majority of individuals of the mapping population, and more than 2 million reads were generated for the parents. Quality analysis showed phred scores of at least 30 (99.9% base-call accuracy) for each base along the length of each read. This high quality gives confidence that each read accurately represents the underlying sequence and detected polymorphisms correspond to actual genotypes. The reads were aligned to the ALB280 (Young et al., 2010) dihaploid reference sequence tags. Approximately 25% of the reads could not be aligned to the reference sequences and were removed from further analysis. A small fraction of the reads (2.5%) aligned with equal probability to multiple sequences. In these instances, only one alignment was used. Initially, 195,430 polymorphic SNP and InDels were identified. Filtering was performed to identify the informative and high-quality polymorphisms. Filtering steps consisted of (i) removal of loci with a samtools quality score less than 20 (Q < 20), (ii) removal of individual genotypes with genotype quality (GQ) scores less than 20 (GQ < 20), (iii) removal of loci that were missing greater than 46 individual genotypes (>20% missing), (iv) removal of loci with more than two alleles, and (v) removal of loci not detected in either of the parents. Finally, adjacent loci that were identified on the same 92-bp ALB280 reference tag were (for our purposes) considered genetically identical and only one locus with the highest quality score was used for further analysis. These steps resulted in 30,020 loci (29,420 SNPs and 600 InDels) that were used for subsequent analysis.

Loci Analysis

Approximately 20% of the filtered loci (4826 SNPs and 134 InDels) displayed a heterozygous genotype in one parent and a homozygous reference genotype in the other parent. Of these loci, approximately 60% (3128; 1012 in female and 2116 in male) segregated with the expected 1:1 ratio in the F_1 population by χ^2 test ($\alpha = 0.05$). To determine if deviation from a 1:1 segregation ratio (distortion) is correlated with any specific properties of the loci, they were separated into nondistorted (SDA) and distorted (TRD-SDA) groups and compared with the nonparametric two-sample Wilcoxon Rank Sum Test (Mann–Whitney) for independence. The comparison results are shown in Supplemental Table S1. Overall, the loci average (median) depth of coverage is 51 (43) and only 4% of the genotypes have a depth of less than 10 reads. The SDA loci average (median) depth of coverage of 48 (43) is significantly less than the TRD-SDA average (median) depth of coverage of 57 (44) ($p < 1 \times 10^{-15}$). Loci positions along the reference tag (and read position) are equivalent between SDA and TRD-SDA ($p = 0.48$). Additionally, an equivalent number of BWA unmapped reads BLAST align to SDA and TRD-SDA reference tags ($p = 0.9$).

Analysis of the individual allele read depth shows that samtools calls a heterozygous genotype when the loci alternate allele read frequency is approximately 15 to 85% of the total and TRD-SDAs have an over- or under-abundance of homozygous F_1 individuals with 95 to 100% reference allele read frequencies. The SDA average (median) alternate allele read frequency for the heterozygous F_1 individuals of 0.47 (0.46) was significantly greater than the TRD-SDA average (median) values of 0.39 (0.36) ($p < 1 \times 10^{-15}$, Supplemental Table S1). The SDA heterozygous parent genotypes have an average (median) alternate allele read frequency equal to the F_1 population (0.47 [0.46]). The distributions of the SDA and TRD-SDA heterozygous parent alternate allele read frequencies are shown in Supplemental Table S2. The TRD-SDA heterozygous parent genotypes have a significantly smaller average (median) alternate allele frequency of 0.42 (0.39) ($p < 1 \times 10^{-15}$). Furthermore, the total number of reads for each heterozygous parent genotype were different with average (median) values of 61 (56) and 70 (59) for SDA and TRD-SDA, respectively ($p < 1 \times 10^{-5}$). In the homozygous reference parent, the reference allele frequencies were determined to be significantly different ($p < 1 \times 10^{-8}$); however, since the data distribution was greatly skewed (88% of the genotypes were 1), this test result was not trusted. The total number of reads for the homozygous reference parent genotypes was also different with average (median) values of 54 (48) and 67 (52) for SDA and TRD-SDA, respectively ($p < 1 \times 10^{-5}$).

These GBS loci were integrated with SSR markers (Okada et al., 2010) and input into JoinMap4 in three parts:

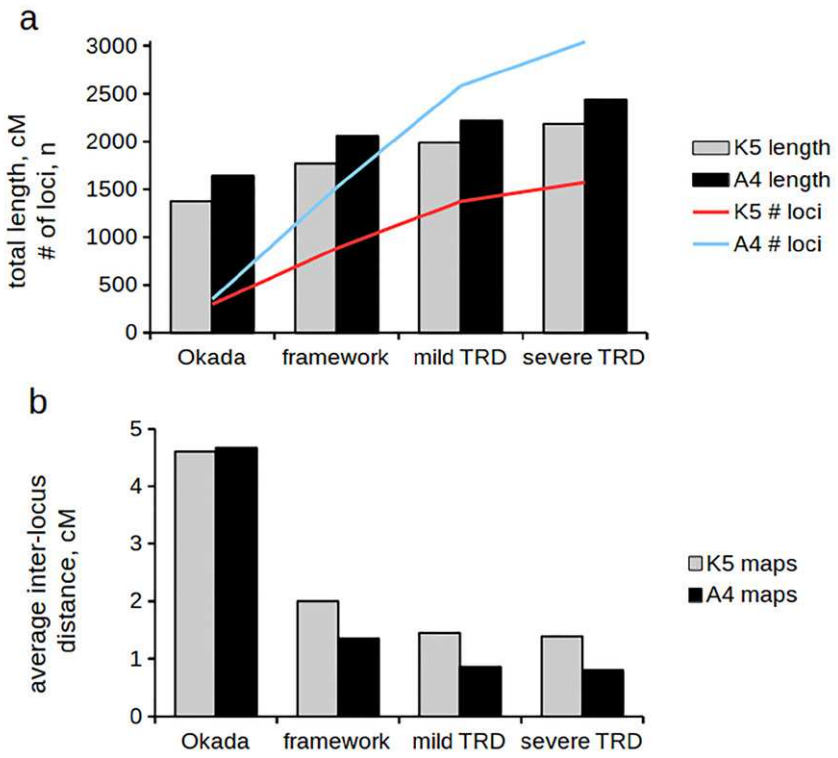

Figure 1. Graphical comparison of summary statistics (Table 1) of different maps of K5 × A4 population. (a) Total length and total number markers in each map. (b) The average interlocus distance (total length/total number of markers).

SDA, mild TRD-SDA, and severe TRD-SDA as determined by severity of transmission segregation distortion (see Materials and Methods). The remaining loci with other parent genotype combinations were not used in mapping.

Linkage Maps

Ninety-nine percent of the SDAs coalesced into 18 LGs at a minimum independence test LOD of 10.0 in both female and male maps. The LGs were named by identification of shared markers between these GBS groups and prior SSR LGs (Okada et al., 2010), which correspond to syntenic diploid chromosomes of foxtail millet with arbitrary "a" and "b" subgenome designations. Upon ML ordering of the LGs, 124 markers were identified as accessory from 884 SDA markers in the female maps and 345 markers were removed as accessory from the 1519 SDA makers in the male maps. Graphical representations of the maternal (K5) and paternal (A4) maps are shown in Fig. 1. A summary of the map statistics is shown in Table 1, graphically displayed in Fig. 2, and individual LG map statistics are shown in Supplemental Table S3. The SDA maps had a total length of 1770 cM and 2059 cM, an average interlocus distances of 2.0 and 1.4, and 14 and 8 gaps over 10 cM for the female and male maps, respectively. Total map lengths are increased 29 and 25% over the previously published female and male framework maps, but marker density significantly increased (Table 1).

Comparison of previously published SSR marker maps (Okada et al., 2010) with GBS maps created here show that the maps are largely collinear, with approximately 90% of the shared markers displaying the same order. Analysis of the shared marker interval distances between the two maps show a median increase of 4 and 11%, with 31 and 39% of intervals increasing over 1 cM for female and male maps, respectively. Comparisons of SSR, GBS, and physical maps are displayed in Table 2. As a metric to determine map marker quality, we used the probability of a correct placement averaged among all the individuals that JoinMap4 returns after regression ordering (–LogP). Only four markers in the female maps and three markers in male maps possessed –LogP values corresponding to less than a 10% chance of correct marker placement.

Segregation Distortion Maps

Two different TRD maps were created to identify regions of segregation distortion, to understand how distortion affects maps and to increase the marker density. Severe TRD maps incorporated all available loci, and mild TRD maps incorporated loci that only moderately deviated from the expected transmission segregation ratio (see Materials an Methods). Mild TRD maps contained 55 and 70% more markers than framework maps for the female and male maps, respectively (not including accessory loci). Severe TRD maps contained 78 and 100% more markers than framework maps for the female and male maps, respectively (Table 1). Markers displaying poor placement (<10% chance of correct placement) after ordering were designated accessory and removed from the TRD map. Many of the markers removed in this way suffered from extreme segregation distortion

Table 1. Summary of paternal genetic linkage maps produced here and traditional marker maps.

	Okada et al., 2010		Serba et al., 2013		Li et al., 2014		Framework		Mild TRD		Severe TRD	
	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male
Parent	Kanlow (K5)	Alamo (A4)	Alamo (AP13)	Summer (VS16)	Alamo/Kanlow (PV281)	Summer/Alamo (NF472)	Kanlow (K5)	Alamo (A4)	Kanlow (K5)	Alamo (A4)	Kanlow (K5)	Alamo (A4)
Map length, cM	1377	1644	1733	1508†	1629	1462	1770	2059	1990	2217	2185	2439
No. mapped markers, *n*	299	352	515	363	445	417	884	1519	1374	2581	1571	3040
No. accessory markers, *n*	264	190	–	–	82	51	134	345	18	40	131	242
Average no. loci per linkage group, *n*	17	20	29	21	25	23	49	84	76	143	87	169
Average interlocus distance, cM	4.6	4.7	3.4	4.2	3.7	3.5	2.0	1.4	1.4	0.9	1.4	0.8

† Seventeen linkage groups (other maps contain 18 linkage groups).

(χ^2 test, $p < 1 \times 10^{-4}$). In the mild TRD maps, 21% of the mapped markers were TRD-SDAs and in the severe TRD maps, 33% of mapped markers were TRD-SDAs; whereas, 42% of all the markers input into JoinMap4 for analysis were TRD-SDAs. Comparison of the total map lengths show an increase of ~10% for the mild TRD maps and an increase of ~20% for the severe TRD maps over the framework maps (Table 2). Direct visualization of the framework and severe TRD maps in Fig. 3 shows that much of this total length increase is due to distorted markers positioned at the ends of many LGs. If these markers are removed (excluding entire arms of Ia-m & IIb-m; 91 and 130 loci removed from female and male severe TRD maps), the total length is equivalent to the total framework map size. Analysis of the shared markers between framework and mild TRD maps show a median interval increase of 1 and 14%, 17 and 23% of intervals increasing over 1 cM, and 81 and 77% of total shared markers retaining orientation for female and male maps, respectively (Table 2). When shared markers between the framework and severe TRD maps are compared, the differences between the maps become greater. For the female map, this resulted in a median marker interval increase of 4%, with 18% of intervals increasing over 1 cM, and 84% of total shared markers retaining orientation. The male map was more affected with a median marker interval increase of 14%, with 27% of intervals increasing over 1 cM, and 69% of total shared markers retaining orientation (Table 2).

Almost every LG possesses at least one region of segregation distortion in the mild and severe TRD maps. These regions are highlighted in Fig. 4. All of the distorted regions identified in the mild TRD maps were also identified in the severe TRD maps, with additional TRD-SDAs generally increasing the size and density of the region. In some cases, half of a LG (Ia-m, Ib-m IIa-m, and IIb-m) or the majority of the markers on a LG (VIIa-m and VIb-f) are comprised of distorted markers in the severe TRD maps. In other cases, pericentromeric or telomeric regions exhibit TRD. Four of these distorted regions displayed high degrees of linkage with each other. The distorted arm of Ia-m is significantly linked to the distorted arm of Ib-m and the distorted end of VIIb-m is significantly linked to the centromeric distorted region of Va-m (Supplemental Figure S2).

Comparison to Physical Maps

The switchgrass draft genome assembly was recently made available (*P. virgatum* v1.1, DOE-JGI, http::://www.phytozome.net/panicumvirgatum) (Goodstein et al., 2012). This assembly was anchored using linkage maps produced from a cross between the genotype AP13 ('Alamo') and genotype VS16 ('Summer'). Approximately half of the assembled contigs have been scaffolded together to create chromosome pseudomolecules. As an orthologous metric of our linkage map quality, we aligned the distorted GBS linkage map markers with the

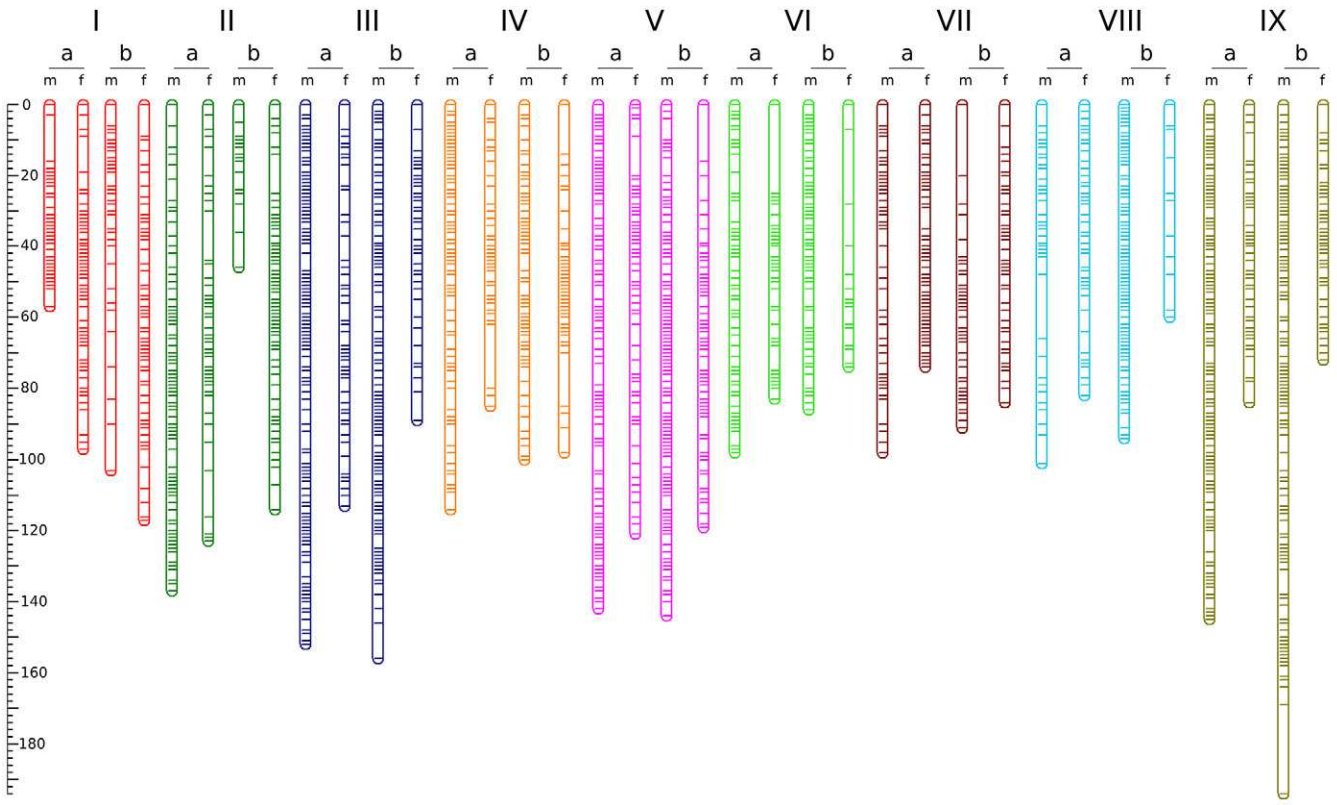

Figure 2. Framework genotyping-by-sequencing male (m) and female (f) linkage maps. Length scale in cM.

Table 2. Comparison of shared markers between genetic and physical maps.

	Comparison between		No. shared markers, *n*	Total length increase	Median shared marker interval increase	LCS[†]
				%		
K5	Okada	Framework	227	29	4	90
	Framework	Mild TRD[‡]	866	12	1	81
	Framework	Severe TRD	846	23	4	84
	Mild TRD	Physical v1.1	714	–	–	55
	Severe TRD	Physical v1.1	861	–	–	50
A4	Okada	Framework	247	25	11	91
	Framework	Mild TRD	1481	8	14	77
	Framework	severe TRD	1455	18	18	69
	Mild TRD	Physical v1.1	1541	–	–	46
	Severe TRD	Physical v1.1	1858	–	–	42

[†] LCS, longest common subsequence as a percentage of the number of shared markers. This value describes the number of shared markers that possess the same order.

[‡] TRD, transmission ratio distortion.

draft assembly using BLAST. In general, the most likely alignments of the markers from each LG were on the same pseudomolecule, allowing unambiguous assignment of each LG to a physical assembly chromosome. Approximately, 60% of the markers in each severe TRD map aligned to the same chromosome. The remaining markers aligned to a different pseudomolecule or unassembled contig (~34%), did not align to the assembly (~1%), or were genomic SSR markers without available sequence information (~5%). Linkage groups II-V, VII, and VIII aligned best with their respective (arbitrary) subgenome designation in the draft assembly and LG I, VI, and IX subgenome designations were switched (e.g., LG Ia aligned to Chr01b). The map positions of the severe TRD maps were plotted against their most likely position in the draft genome and the results are shown in Fig. 4. The framework and mild TRD map positions were also plotted against their most likely alignments to the draft genome and are shown in Supplemental Figures S2 and S3. At the chromosome-length scale, strong collinearity exists between the genetic and physical maps. However, a significant number of differences exist at short-length scales, especially in the pericentromeric regions. This local misordering is reflected in ~50% order retention between the TRD and physical maps. Interestingly, a few of the shorter LGs appear to only contain markers from one arm of the chromosome.

Discussion

Genotyping-by-Sequencing Linkage Maps

Switchgrass linkage maps were produced using a full-sib population derived from a cross between two heterozygous genotypes sampled from lowland tetraploid cultivars Kanlow (K5) and Alamo (A4) as female and male, respectively. Each of the parent maps contained a complete representation of 18 LGs in which the expected nine pairs of homeologues were identified. Incorporation of markers exhibiting transmission segregation distortion further

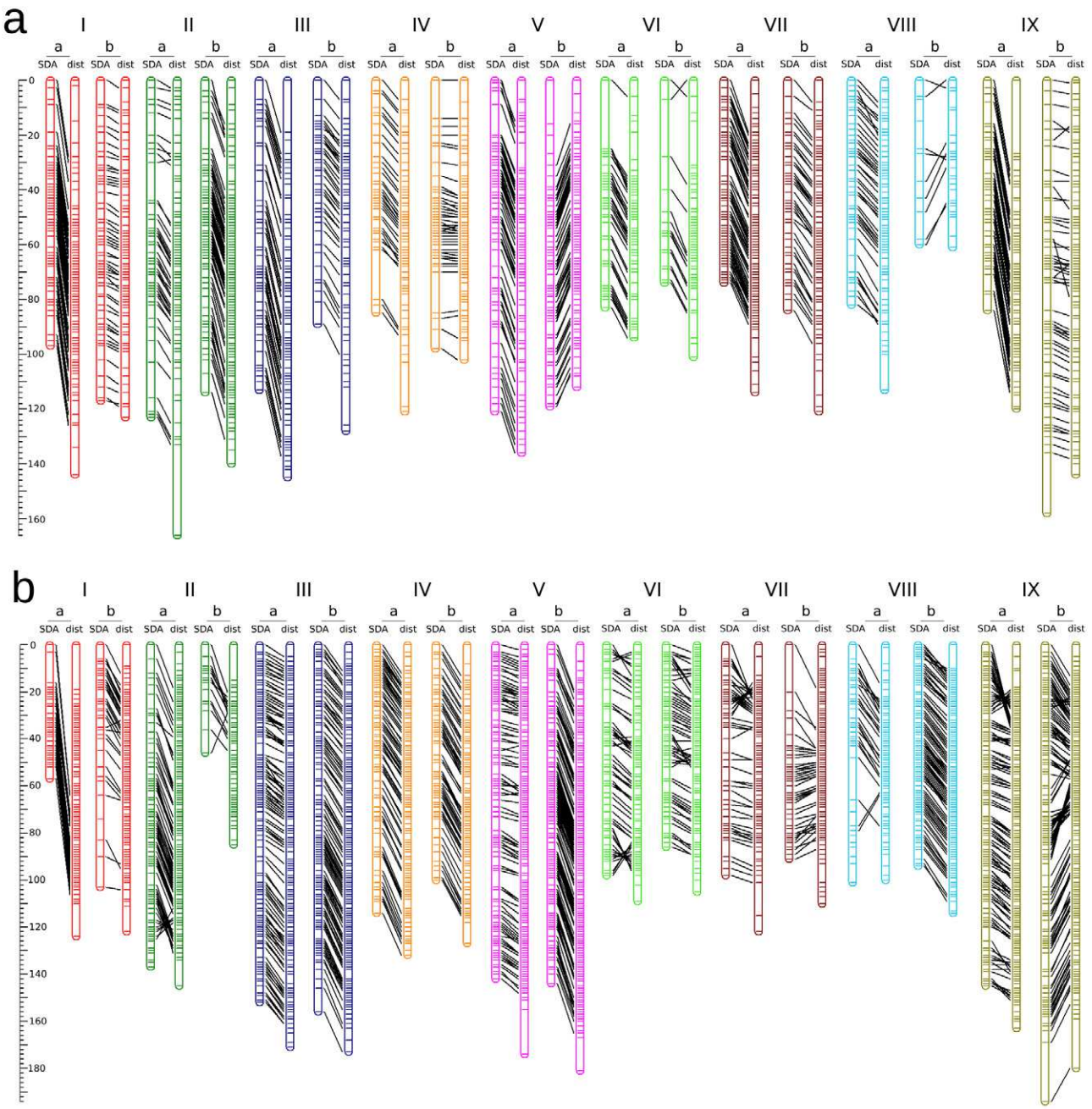

Figure 3. Comparison of framework (single-dose allele [SDA]) maps with severe transmission ratio distortion (dist) maps for the (a) female and (b) male maps. Black lines connect shared markers on the two maps and qualitatively highlight collinearity. Length scale in cM.

increased the density of the map and enabled an analysis of genome structure and comparison to the draft genome assembly.

Genotyping-by-sequencing was used to rapidly genotype the 231 progeny and parents of this mapping cross. From the sequencing capacity of less than three lanes of one Illumina Hi-Seq flow-cell, 350 million reads were obtained, and approximately 5000 high-quality markers were identified. This number of markers were adequate for producing the high-density linkage maps and was likely limited by the degree of heterozygosity of the parental lines and the choice of restriction enzyme for the library preparation.

Male vs. Female Maps

Although similar numbers of sequencing reads were generated for both parents, the male linkage maps contained approximately double the number of informative markers than the female maps. Analysis of the parent genotypes revealed that approximately 3000 more heterozygous SNPs were identified before filtering out polymorphisms with missing F_1 genotypes. This suggests that the Kanlow parent possessed less nucleotide diversity than the Alamo cultivar. This could be due to the histories of these two populations. Kanlow underwent breeding at the Manhattan Plant Materials Center whereas the Alamo individual originated from an environmental accession (Hultquist et al., 1997).

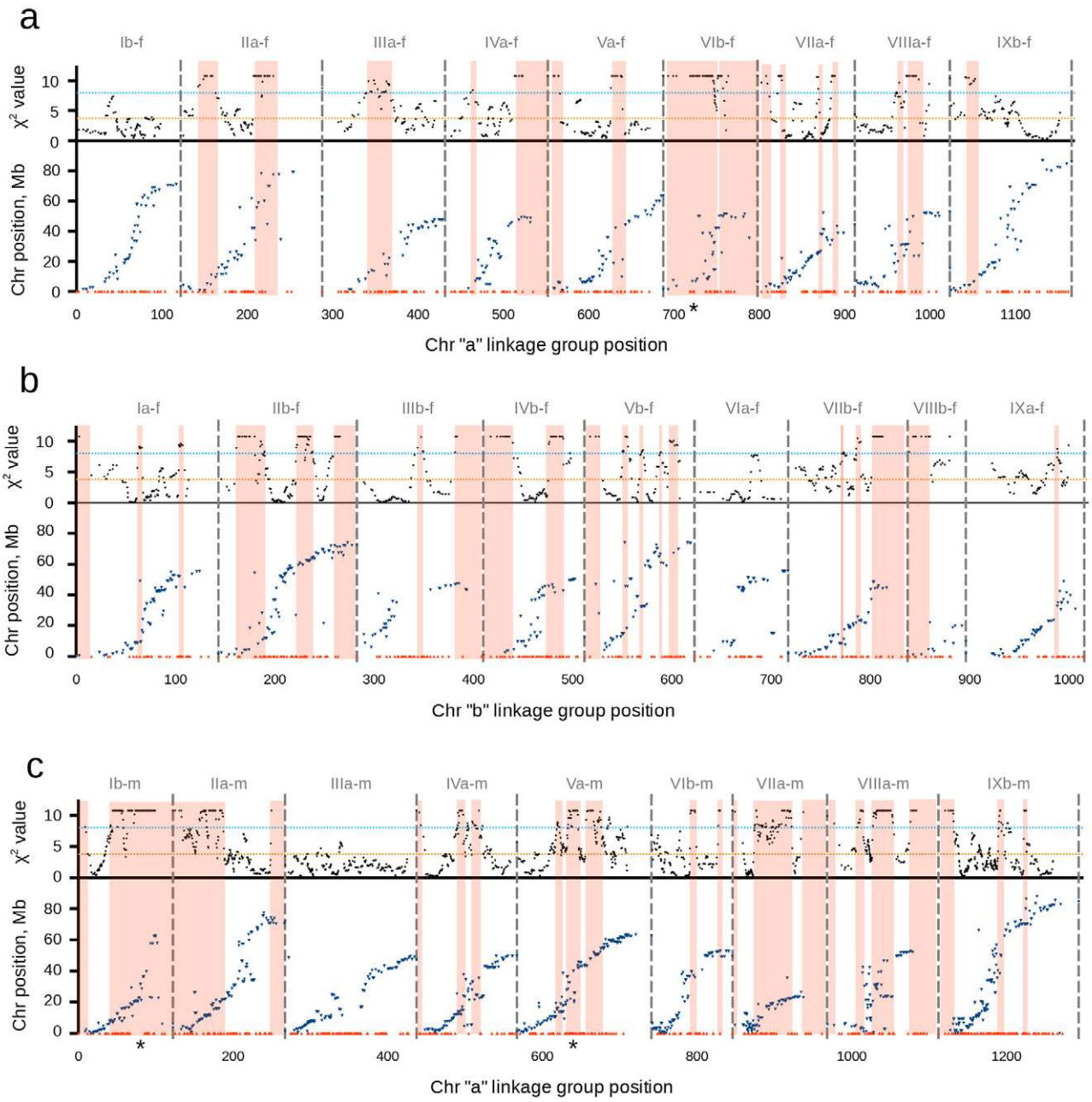

Figure 4. Plot of severe transmission ratio distortion (TRD) map with physical map. χ^2 value and BLASTn aligned position on the v1.1 *Panicum virgatum* genome assembly is plotted vs. the concatenated severe TRD map for female linkage groups (LGs) aligning to the (a) "a" homeolog; (b) "b" homeolog; male LGs aligning to the (c) "a" homeolog and (d) "b" homeolog. Top pane: χ^2 value testing for deviation from 1:1 segregation ratio. Orange and blue lines indicate χ^2 significance values of $p = 0.05$ and $p = 0.01$, respectively. Black data points indicate sliding-scale average χ^2 value (capped at $\chi^2 = 10$ for clarity). Bottom pane: Physical map alignment. Blue data points indicate physical map position and red data points along the *x*-axis indicate markers on the genetic map that did not align to a position on the physical map. Highlighted regions indicate areas of contiguous TRD and asterisk (*) denotes regions of segregation distortion reported previously (Okada et al., 2010).

Comparison to Published Maps

The framework linkage maps generated here used GBS markers integrated with PCR-based SSR markers previously used to map the switchgrass genome (Okada et al., 2010). This integration allowed a direct comparison to the previously published linkage maps. This comparison highlights a number of key points. First, the shared markers between the two sets of maps are largely collinear, with less than 8% inversion in the relative orders and only small increases in the shared marker intervals. (Table 2) Second, the GBS maps are much more dense than prior SSR maps as the number of potential markers greatly increased with the switch to in silico genotyping. Third, even with stringent QC of GBS markers, the maps generated here are larger than genetic maps previously published (Okada et al., 2010; Liu et al.,

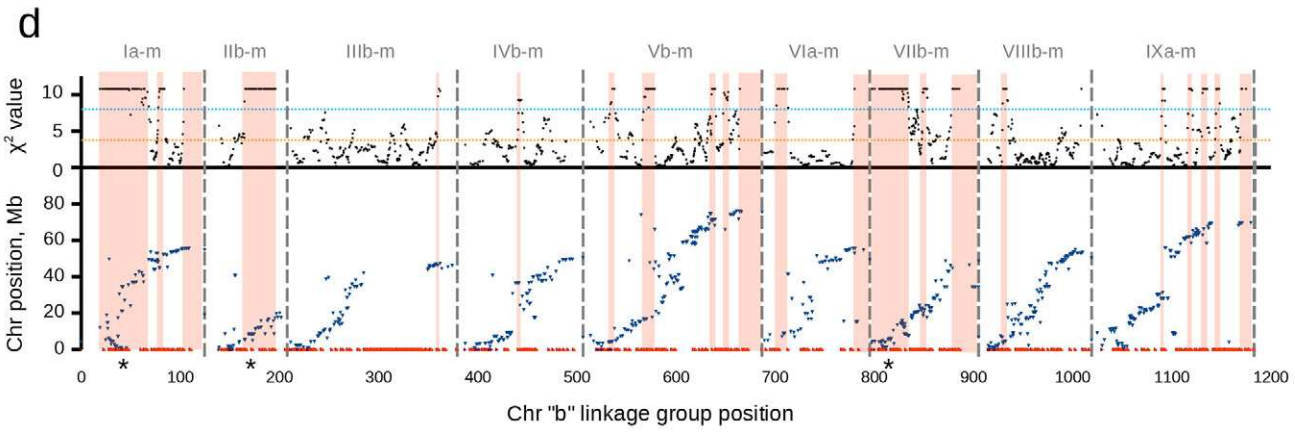

Figure 4. Continued.

2012; Serba et al., 2013; Li et al., 2014). Total map lengths were increased by over 25%, but analysis of the common marker intervals shows a median increase of much less than 25% (Table 2). In addition, the GBS LGs that exhibited the greatest length increases (>70%) over the previous LGs contained very few markers (<14) (Supplemental Table S3). In other grass crops, such as wheat (*Triticum aestivum* L.) and rice (*Oryza sativa* L.), EST-SSR markers did not evenly cover the entire genome (Wu et al., 2002; Yu et al., 2004). In sorghum [*Sorghum bicolor* (L.) Moench], genomic SSR markers poorly identified 25% of the genome (Bhattramakki et al., 2000). The increased map density suggests that we have identified GBS markers in many regions of the switchgrass genome lacking prior SSR marker coverage and some of the increase in total map size is due to the addition of these unidentified regions. However, this may not entirely explain the size, as the framework maps are also larger than other traditional marker switchgrass linkage maps generated from the AP13 × VS16 population (Table 1) (Serba et al., 2013; Li et al., 2014). One additional report used GBS protocols to identify a very large number of switchgrass SNPs (Lu et al., 2013). These loci were grouped into 18 LGs, but were ordered using syntentic chromosomes of foxtail millet and the total LG size was not reported. Linkage maps generated with GBS data in other plants, such as red raspberry (*Rubus idaeus* L.) (Ward et al., 2013) and rice (Spindel et al., 2013), have shown total map length increases compared with maps generated with previous PCR-based markers. These increases may be due to inherent shortcomings of GBS protocols that result in poor-quality genotypes and missing data. These factors lead to incorrect recombination fraction estimates and incorrect genetic distances between markers (Hackett and Broadfoot, 2003). The JoinMap4 regression algorithm performs relatively well with regard to missing data and incorrect genotypes in comparison to other mapping software (Wu et al., 2011), but it does not handle large data sets efficiently. Although not investigated in this report, implementation of newer packages such as OneMap (Margarido et al., 2007) or Tmap (Cartwright et al., 2007) may be better suited to handle the large number of GBS genotypes that will be common in the future (Cheema and Dicks, 2009).

Transmission Distortion

Markers that deviate from the expected Mendelian inheritance ratios are commonly identified in genetic mapping studies and are caused by a variety of processes that can influence viability at any developmental stage from gamete formation to zygote development (Zamir and Tadmor, 1986; Lyttle, 1991). Reduced recombinations around centromeres and telemeric effects can also cause apparent segregation distortion (Debener and Mattiesch, 1999). These distorted markers can introduce inaccuracies to maps due to spurious linkages, biased estimates of recombination fractions, and incorrect marker order (Lorieux et al., 1995; Cloutier et al., 1997).

To determine if the segregation distortion is due to biases in the GBS pipeline, we analyzed various properties of the GBS loci. The removal of low-quality genotypes during filtering did not artificially skew the F_1 inheritance ratio, as inclusion of genotypes with samtools GQ scores <20 resulted in a net of 142 more TRD-SDA by χ^2 test ($\alpha = 0.05$). The BWA alignment step also did not contribute to distortion, as the unmapped reads aligned to SDA and TRD-SDAs reference tags similarly. Position along the read (reference tag) also did not correlate with distortion, as SDAs and TRD-SDAs were distributed evenly on the read. Undercalling of F_1 heterozygotes (which is common in low-coverage techniques) was not a factor in this study as the average depth of coverage was >40 for every locus (Supplemental Table S1).

With highly repetitive sequences, paralogous sites may collapse onto a single reference tag on alignment. For example, approximately 6% of the parent genotypes called as heterozygotes were likely artifacts of this collapsing property as judged by the lack of homozygotes among the progeny (none of these were mapped). Increased depth of coverage and decreased alternate allele read frequency in the plants correlated with segregation distortion, which suggests that multiple site collapses might cause apparent segregation distortion. However, similar numbers of SDAs and TRD-SDAs are

observed at every heterozygous parent alternate allele read frequency (Supplemental Table S2) and less than expected alternate allele read frequencies are associated with TRD-SDAs inheritance ratios skewed to more *and* less heterozygous F_1 progeny than expected.

Linkage of the collapsing paralogous sequences, and presence or absence of other alleles, could also produce incorrect genotypes that skew heritability. To further shed light on this, we analyzed 618 pairs of SNPs that occurred on the same 92-base reference tag. In 505 cases, only two haplotypes were detected: 282 pairs were SDA and 223 pairs were TRD-SDA. This ratio of SDA to TRD-SDA is approximately the same as the entire GBS data set. The remaining cases possessed more than two haplotypes and were a result of multiple site alignment. Although many loci appear to collapse onto single reference tags, this property of the genotyping pipeline is not likely a primary cause of apparent segregation distortion, and many TRD-SDAs were placed on the linkage map with high confidence and displayed good collinearity with the v1.1 physical map.

Incomplete disomic inheritance of the homeologous switchgrass chromosomes may also account for some of the apparent segregation distortion. However, only 6% of the mapped severe TRD-SDAs fit in this category by χ^2 test (3:1 ratio; $\alpha = 0.05$). These markers are present on most of the severe TRD maps and appear to cluster only in distorted regions of Ia, Ib, IIb, and VIIb in the male maps. Since previous studies did not find significant evidence of tetrasomic inheritance in switchgrass (Missaoui et al., 2005; Okada et al., 2010) and we only identified a small fraction of GBS markers that segregate nondisomically (and do not cluster on all the homeologs), this is also not likely a primary cause of segregation distortion in this study.

Comparison of mild and severe TRD maps with framework maps show different effects in the female and male subgenomes. For the female maps, the mild TRDs map lengths increased 12% over the framework, and the severe TRD map lengths increased 23% over the framework. Although the total lengths were increased by a significant amount, the median shared-marker interval increased less than 4% (Table 2) and visual comparison of the maps show distorted markers at the ends of many of the LGs (Fig. 3a). This suggests that these end-located TRD-SDAs likely are poorly linked with the majority of the LGs and these regions may not be a realistic representation of the genome. In addition, we also found that 12 and 17% of the shared markers between the framework and mild and severe TRD maps exhibit an inverted order. This is likely due to linkage of SDAs with TRD-SDAs that are poorly placed. The male maps show less of an increase in total map length between framework and TRD maps than the female maps, (increase of 8 and 18% for mild and severe TRD maps, Table 2). Like the female maps, many of these TRD markers locate to the end of the LGs (Fig. 3b), but unlike female maps, the shared marker intervals increased by a significant amount with median values of 14 and 18% for mild and severe TRD maps, respectively (Table 2). The total length increase is likely due to end-located TRD-SDAs because removal of these regions decrease the total map size to the same as the framework maps. The relatively large increases in shared marker intervals are likely a result of more rearrangements of the interstitial SDA markers due to linkage with poorly placed distorted regions. This is visually observed in Fig. 3b and quantified with the large number of shared marker inversions that occur (22 and 27% in mild and severe TRD maps, Table 2). Distorted markers undoubtedly cause more adverse effect in the male TRD maps. Considering the nature of this cross, this analysis suggests that some of the segregation distortion and multilocus interactions observed in the male map are due to pollen-specific gametophytic incompatibility effects (Li et al., 1997; Martínez-Reyna and Vogel, 2002; Shinozuka et al., 2010), which eliminate male gametes with specific allele combinations.

Clustering of distorted markers can help identify distorter loci present due to lethal gene alleles or incompatibility loci (Lefebvre et al., 1995). These distorter loci are important factors in sexual reproduction (Lu et al., 2002; Shinozuka et al., 2010) and are useful for advanced breeding efforts, although precise mapping of these regions is difficult (Jenczewski et al., 1997; Lu et al., 2002; McDaniel et al., 2007). In the TRD maps, many TRD-SDAs clustered together. All distorted regions identified in previously published maps (Okada et al., 2010) were also observed in the TRD maps along with the clustering of additional TRD-SDAs and provides greater confidence of a legitimate distorter loci near those regions. Map regions that contain self-incompatibility loci are often distorted and exhibit interactions with other LGs in perennial ryegrass (*Lolium perenne* L.) mapping populations (Thorogood et al., 2002; Hackauf and Wehling, 2005; Klaas et al., 2011). The distorted region on Va-m near the centromere was significantly linked with the telomeric distorted region on VIIb-m. This distorted region of Va-m is syntenic with regions on related grass genomes that contain the *Z*-locus of the two-component self-incompatibility system (Shinozuka et al., 2010). Interestingly, synteny also predicts that the complement self-incompatibility *S*-locus is on switchgrass LG III (Okada et al., 2010) and not on the linked distorted region of VIIb-m. Many TRD-SDAs are clustered together near the centromere and telomeres of most of the LGs and are likely distorted due to the lack of recombination near the centromere and telomeric effects (Debener and Mattiesch, 1999). Alternatively, some distorted regions located at the ends of LGs also increase the size of the maps. In these cases, these regions are likely a result of the JoinMap4 regression ordering algorithm creating artificial genetic distance between the LGs and poorly linked distorted regions. The remaining interstitial distorted clusters may be indicative of an underlying distorter loci, but there appears to be more regions than would be expected. Some of these regions may simply be incorrectly placed, as segregation distortion causes problems with estimating recombination fractions. The

TRD-SDAs possessed smaller JoinMap4 probability values that corresponded to less likely map placement than nondistorted loci (Supplemental Table S1). Indeed, some of the largest interval differences of shared markers between framework and TRD maps appear to be caused by groups of TRD-SDAs interfering with SDAs. Segregation distortion has not been thoroughly analyzed in the switchgrass genome and it will take more genetic resources to identify authentic distorter loci, to accurately position TRD-SDAs, and to understand the impact that multiple site collapse has on GBS genotyping.

Physical Map Comparison

More than half of the mapped markers from the mild and severe TRD maps aligned to the draft genome assembly, with markers from each LG primarily aligning to one pseudomolecule. The plots created with framework (Supplemental Figure S2), mild TRD (Supplemental Figure S3), and severe TRD (Fig. 4) maps look very similar except for the higher marker density and larger distorted regions on the severe TRD map. These plots show global collinearity of the physical and genetic map; however, many instances of nonlinearity occur at smaller length scales.

Longest common subsequence analysis of the order of markers that align to the physical assembly and mild and severe TRD genetics maps suggests only moderate collinearity with ~50% the shared markers retaining the same order on both maps (Table 2). However, most of this misordering is likely due to small-length scale inversion that skews the LCS values. The sigmodial-like curve of the plots suggests less recombination proximal to the metacentric centromeres (Young et al., 2012). Interestingly, LGs IIb-m, VIIa-m, and VIIIb-f are much shorter then the remaining LGs; IIb-m and VIIa-m appear to have many markers localized only to one arm of the chromosome. The draft assembly contains sufficient genetic information in these pseudomolecules (2b and 7a), as evidenced by alignment with these LGs in the female linkage maps. Therefore, these small sizes are likely due to lack of heterozygous loci in the missing arms in the male LGs.

This global collinearity of the alignment also provides an additional level of confidence in the quality of the map as the assembly was anchored by genetic maps generated with different parents (Serba et al., 2013). Genotyping-by-sequencing marker systems are typically unique for restriction enzyme combination, and comparisons between different maps can be difficult. In this case, the GBS maps can be indirectly compared with VS16 × AP13 maps via the genome assembly and to previously published K5 × A4 maps via shared SSR markers (Okada et al., 2010). Approximately 30% of the mapped markers aligned to contigs not scaffolded into pseudomolecules in the draft assembly and half of the mapped markers differ in orientation at small-length scales. This information will be valuable for improving future switchgrass genome assemblies, as will accessory markers that have not been ordered relative to nearby markers but that have assigned map positions.

Here we present a high-density genetic linkage map from a population derived from two lowland switchgrass germplasms. Genotyping-by-sequencing was used to rapidly identify many novel markers, and by integrating these with previous PCR-based markers, we can directly compare maps generated with different genotyping technologies. The GBS maps are more informative due to the increase in marker density and the inclusion of markers exhibiting transmission segregation distortion. The GBS maps populate regions not previously mapped, and several regions of segregation distortion indicate a number of loci impacting gametic or postzygotic fitness. These maps will be useful to help refine the physical genome assembly and protocols developed here will be used for further analysis using GBS switchgrass germplasm. As a community resource, the linkage maps, population genotypes, and sequence information will be made available for download at http://wheat.pw.usda.gov/pubs/2015/Fiedler.

Acknowledgments

This work was supported by the USDA–ARS, CRIS 9235-21000-017-00D, National Program 301, Plant Genetic Resources, Genomics and Genetic Improvement. The work conducted by the US Department of Energy Joint Genome Institute was supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. The draft genome sequence data were produced by the US Department of Energy Joint Genome Institute. The USDA–ARS is an equal opportunity/affirmative action employer, and all agency services are available without discrimination. Mention of commercial products and organizations in this manuscript is solely to provide specific information. It does not constitute endorsement by USDA–ARS over other products and organizations not mentioned.

References

Baird, N.A., P.D. Etter, T.S. Atwood, M.C. Currey, A.L. Shiver, Z.A. Lewis, E.U. Selker, W.A. Cresko, and E.A. Johnson. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3:E3376. doi:10.1371/journal.pone.0003376

Bauer, D.F. 1972. Constructing confidence sets using rank statistics. J. Am. Stat. Assoc. 67:687–690. doi:10.1080/01621459.1972.10481279

Bhandari, H.S., M.C. Saha, P.N. Mascia, V.A. Fasoula, and J.H. Bouton. 2010. Variation among half-sib families and heritability for biomass yield and other traits in lowland switchgrass (L.). Crop Sci. 50:2355–2363. doi:10.2135/cropsci2010.02.0109

Bhattramakki, D., J. Dong, A.K. Chhabra, and G.E. Hart. 2000. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.). Moench. Genome 43:988–1002. doi:10.1139/gen-43-6-988

Carroll, A., and C. Somerville. 2009. Cellulosic biofuels. Annu. Rev. Plant Biol. 60:165–182. doi:10.1146/annurev.arplant.043008.092125

Cartwright, D.A., M. Troggio, R. Velasco, and A. Gutin. 2007. Genetic mapping in the presence of genotyping errors. Genetics 176:2521–2527. doi:10.1534/genetics.106.063982

Casler, M.D., C.M. Tobias, S.M. Kaeppler, C.R. Buell, Z.Y. Wang, P. Cao, J. Schmutz, and P. Ronald. 2011. The Switchgrass genome: Tools and Strategies. Plant Gen. 4:273–282. doi:10.3835/plantgenome2011.10.0026

Casler, M.D., K.P. Vogel, C.M. Taliaferro, N.J. Ehlke, J.D. Berdahl, E.C. Brummer, R.L. Kallenbach, C.P. West, and R.B. Mitchell. 2007. Latitudinal and longitudinal adaptation of switchgrass populations. Crop Sci. 47:2249–2260. doi:10.2135/cropsci2006.12.0780

Chagné, D., K. Gasic, R.N. Crowhurst, Y. Han, H.C. Bassett, D.R. Bowatte, T.J. Lawrence, E.H.A. Rikkerink, S.E. Gardiner, and S.S. Korban. 2008. Development of a set of SNP markers present in expressed genes of the apple. Genomics 92:353–358. doi:10.1016/j.ygeno.2008.07.008

Cheema, J., and J. Dicks. 2009. Computational approaches and software tools for genetic linkage map estimation in plants. Brief. Bioinform. 10:595–608. doi:10.1093/bib/bbp045

Chen, D.H., and P.C. Ronald. 1999. A rapid DNA Minipreparation method suitable for AFLP and other PCR applications. Plant Mol. Biol. Rep. 17:53–57. doi:10.1023/A:1007585532036

Cloutier, S., M. Cappadocia, and B.S. Landry. 1997. Analysis of RFLP mapping inaccuracy in *Brassica napus* L. TAG. Theor. Appl. Genet. 95:83–91. doi:10.1007/s001220050535

Colasuonno, P., M.A. Maria, A. Blanco, and A. Gadaleta. 2013. Description of durum wheat linkage map and comparative sequence analysis of wheat mapped DArT markers with rice and Brachypodium genomes. BMC Genet. 14:114. doi:10.1186/1471-2156-14-114

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. Lond. B Biol. Sci. 363:557–572. doi:10.1098/rstb.2007.2170

Danecek, P., A. Auton, G. Abecasis, C.A. Albers, E. Banks, M.A. DePristo, R.E. Handsaker, G. Lunter, G.T. Marth, S.T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158. doi:10.1093/bioinformatics/btr330

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nat. Rev. Genet. 12:499–510. doi:10.1038/nrg3012

Debener, T., and L. Mattiesch. 1999. Construction of a genetic linkage map for roses using RAPD and AFLP markers. Theor. Appl. Genet. 99:891–899. doi:10.1007/s001220051310

Edgar, R.C. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461. doi:10.1093/bioinformatics/btq461

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6:E19379.

Ganal, M.W., A. Polley, E.M. Graner, J. Plieske, R. Wieseke, H. Luerssen, and G. Durstewitz. 2012. Large SNP arrays for genotyping in crop plants. J. Biosci. 37:821–828. doi:10.1007/s12038-012-9225-3

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE 9:E90346. doi:10.1371/journal.pone.0090346

Goodstein, D.M., S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D.S. Rokhsar. 2012. Phytozome: A comparative platform for green plant genomics. Nucleic Acids Res. 40:D1178–D1186. doi:10.1093/nar/gkr944

Grattapaglia, D., and R. Sederoff. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: Mapping strategy and RAPD markers. Genetics 137:1121–1137.

Hackauf, B., and P. Wehling. 2005. Approaching the self-incompatibility locus Z in rye (*Secale cereale* L.) via comparative genetics. Theor. Appl. Genet. 110:832–845. doi:10.1007/s00122-004-1869-4

Hackett, C.A., and L.B. Broadfoot. 2003. Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. Heredity 90:33–38. doi:10.1038/sj.hdy.6800173

Hultquist, S., K. Vogel, D. Lee, K. Arumuganathan, and S. Kaeppler. 1997. DNA content and chloroplast DNA polymorphisms among switchgrasses from remnant Midwestern prairies. Crop Sci. 37:595–598. doi:10.2135/cropsci1997.0011183X003700020047x

Hyten, D.L., S.B. Cannon, Q. Song, N. Weeks, E.W. Fickus, R.C. Shoemaker, J.E. Specht, A.D. Farmer, G.D. May, and P.B. Cregan. 2010. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38. doi:10.1186/1471-2164-11-38

Jachner, S., K.G. Van Den Boogaart, and T. Petzoldt. 2007. Statistical methods for the qualitative assessment of dynamic models with time delay (R Package qualV). J. Stat. Softw. 22:1–30.

Jansen, J., A.G. de Jong, and J.W. van Ooijen. 2001. Constructing dense genetic linkage maps. Theor. Appl. Genet. 102:1113–1122. doi:10.1007/s001220000489

Jenczewski, E., M. Gherardi, I. Bonnin, J.M. Prosperi, I. Olivieri, and T. Huguet. 1997. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). Theor. Appl. Genet. 94:682–691. doi:10.1007/s001220050466

Klaas, M., B. Yang, M. Bosch, D. Thorogood, C. Manzanares, I.P. Armstead, F.C.H. Franklin, and S. Barth. 2011. Progress towards elucidating the mechanisms of self-incompatibility in the grasses: Further insights from studies in Lolium. Ann. Bot. (Lond.) 108:677–685. doi:10.1093/aob/mcr186

Kopecký, D., and B. Studer. 2013. Emerging technologies advancing forage and turf grass genomics. Biotechnol. Adv. 32:190–199. doi:10.1016/j.biotechadv.2013.11.010

Kosambi, D.D. 1943. The estimation of map distances from recombination values. Ann. Eugen. 12:172–175. doi:10.1111/j.1469-1809.1943.tb02321.x

LaFramboise, T. 2009. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. Nucleic Acids Res. 37: 4181–4193.

Lefebvre, V., A. Palloix, C. Caranta, and E. Pochard. 1995. Construction of an intraspecific integrated linkage map of pepper using molecular markers and doubled-haploid progenies. Genome 38:112–121. doi:10.1139/g95-014

Li, H., and R. Durbin. 2010. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26:589–595. doi:10.1093/bioinformatics/btp698

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352

Li, X., N. Paech, J. Nield, D. Hayman, and P. Langridge. 1997. Self-incompatibility in the grasses: Evolutionary relationship of the S gene from *Phalaris coerulescens* to homologous sequences in other grasses. Plant Mol. Biol. 34:223–232. doi:10.1023/A:1005802327900

Li, G., D.D. Serba, M.C. Saha, J.H. Bouton, C.L. Lanzatella, and C.M. Tobias. 2014. Genetic linkage mapping and transmission ratio distortion in a three-generation four-founder population of *Panicum virgatum* (L.). G3 4:913–923. doi:10.1534/g3.113.010165

Liu, L., and Y. Wu. 2012. Identification of a selfing compatible genotype and mode of inheritance in switchgrass. BioEnergy Res. 5:662–668. doi:10.1007/s12155-011-9173-z

Liu, L., Y. Wu, Y. Wang, and T. Samuels. 2012. A high-density simple sequence repeat-based genetic linkage map of switchgrass. G3 2:357–370.

Lorieux, M., B. Goffinet, X. Perrier, D.G. de León, and C. Lanaud. 1995. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross populations. Theor. Appl. Genet. 90:73–80. doi:10.1007/BF00220998

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, E.S. Buckler, and D.E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: Novel insights from a network-based SNP discovery protocol. PLoS Genet. 9(1):E1003215. doi:10.1371/journal.pgen.1003215

Lu, H., J. Romero-Severson, and R. Bernardo. 2002. Chromosomal regions associated with segregation distortion in maize. Theor. Appl. Genet. 105:622–628. doi:10.1007/s00122-002-0970-9

Lyttle, T.W. 1991. Segregation distorters. Annu. Rev. Genet. 25:511–557. doi:10.1146/annurev.ge.25.120191.002455

Mammadov, J., R. Aggarwal, R. Buyyarapu, and S. Kumpatla. 2012. SNP markers and their impact on plant breeding. Int. J. Plant Genomics 2012:728398. doi:10.1155/2012/728398

Margarido, G.R.A., A.P. Souza, and A.A.F. Garcia. 2007. OneMap: Software for genetic mapping in outcrossing species. Hereditas 144:78–79. doi:10.1111/j.2007.0018-0661.02000.x

Martínez-Reyna, J., and K. Vogel. 2002. Incompatibility systems in switchgrass. Crop Sci. 42:1800–1805. doi:10.2135/cropsci2002.1800

McDaniel, S.F., J.H. Willis, and A.J. Shaw. 2007. A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. Genetics 176:2489–2500. doi:10.1534/genetics.107.075424

McLaughlin, S.B., and L. Adams Kszos. 2005. Development of switchgrass (*Panicum virgatum*) as a bioenergy feedstock in the United States. Biomass Bioenergy 28:515–535. doi:10.1016/j.biombioe.2004.05.006

Micheletti, D., M. Troggio, A. Zharkikh, F. Costa, M. Malnoy, R. Velasco, and S. Salvi. 2011. Genetic diversity of the genus *Malus* and implications for linkage mapping with SNPs. Tree Genet. Genomes 7:857–868. doi:10.1007/s11295-011-0380-8

Missaoui, A.M., A.H. Paterson, and J.H. Bouton. 2005. Investigation of genomic organization in switchgrass (*Panicum virgatum* L.) using DNA markers. Theor. Appl. Genet. 110:1372–1383. doi:10.1007/s00122-005-1935-6

Monti, A., G. Bezzi, G. Pritoni, and G. Venturi. 2008. Long-term productivity of lowland and upland switchgrass cytotypes as affected by cutting frequency. Bioresour. Technol. 99:7425–7432. doi:10.1016/j.biortech.2008.02.034

Narasimhamoorthy, B., M. Saha, T. Swaller, and J. Bouton. 2008. Genetic diversity in switchgrass collections assessed by EST-SSR markers. BioEnergy Res. 1:136–146. doi:10.1007/s12155-008-9011-0

Nielsen, R., J.S. Paul, A. Albrechtsen, and Y.S. Song. 2011. Genotype and SNP calling from next-generation sequencing data. Nat. Rev. Genet. 12:443–451. doi:10.1038/nrg2986

Okada, M., C. Lanzatella, M.C. Saha, J. Bouton, R. Wu, and C.M. Tobias. 2010. Complete switchgrass genetic maps reveal subgenome collinearity, preferential pairing and multilocus interactions. Genetics 185:745–760. doi:10.1534/genetics.110.113910

Parrish, D., and J. Fike. 2005. The biology and agronomy of switchgrass for biofuels. Crit. Rev. Plant Sci. 24:423–459. doi:10.1080/07352680500316433

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS ONE 7:E32253. doi:10.1371/journal.pone.0032253

Romay, M.C., M.J. Millard, J.C. Glaubitz, J.A. Peiffer, K.L. Swarts, T.M. Casstevens, R.J. Elshire, C.B. Acharya, S.E. Mitchell, S.A. Flint-Garcia, M.D. McMullen, J.B. Holland, E.S. Buckler, and C.A. Gardner. 2013. Comprehensive genotyping of the USA national maize inbred seed bank. Genome Biol. 14:R55. doi:10.1186/gb-2013-14-6-r55

Schmer, M., K. Vogel, R. Mitchell, and R. Perrin. 2008. Net energy of cellulosic ethanol from switchgrass. Proc. Natl. Acad. Sci. USA 105:464–469. doi:10.1073/pnas.0704767105

Serba, D., L. Wu, G. Daverdin, B.A. Bahri, X. Wang, A. Kilian, J.H. Bouton, E.C. Brummer, M.C. Saha, and K.M. Devos. 2013. Linkage maps of lowland and upland tetraploid switchgrass ecotypes. BioEnergy Res. 6:953–965. doi:10.1007/s12155-013-9315-6

Shinozuka, H., N.O.I. Cogan, K.F. Smith, G.C. Spangenberg, and J.W. Forster. 2010. Fine-scale comparative genetic and physical mapping supports map-based cloning strategies for the self-incompatibility loci of perennial ryegrass (*Lolium perenne* L.). Plant Mol. Biol. 72:343–355. doi:10.1007/s11103-009-9574-y

Spindel, J., M. Wright, C. Chen, J. Cobb, J. Gage, S. Harrington, M. Lorieux, N. Ahmadi, and S. McCouch. 2013. Bridging the genotyping gap: Using genotyping-by-sequencing (GBS) to add high-density SNP markers and new value to traditional bi-parental mapping and breeding populations. Theor. Appl. Genet. 126:2699–2716. doi:10.1007/s00122-013-2166-x

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J. 3:739–744. doi:10.1111/j.1365-313X.1993.00739.x

Thorogood, D., W.J. Kaiser, J.G. Jones, and I. Armstead. 2002. Self-incompatibility in ryegrass 12. Genotyping and mapping the *S* and *Z* loci of *Lolium perenne* L. Heredity 88:385–390. doi:10.1038/sj.hdy.6800071

Tobias, C.M., D.M. Hayden, P. Twigg, and G. Sarath. 2006. Genic microsatellite markers derived from EST sequences of switchgrass (*Panicum virgatum* L.). Mol. Ecol. Notes 6:185–187. doi:10.1111/j.1471-8286.2006.01187.x

Triplett, J.K., Y. Wang, J. Zhong, and E.A. Kellogg. 2012. Five nuclear loci resolve the polyploid history of switchgrass (*Panicum virgatum* L.) and relatives. PLoS One 7: E38702.

van Ooijen, J.W. 2006. Joinmap 4: Software for the calculation of genetic linkage maps in experimental populations. Plant Research International B.V., Wageningen, Netherlands.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34:275–305.

Voorrips, R.E. 2002. MapChart: Software for the graphical presentation of linkage maps and QTLs. J. Hered. 93:77–78. doi:10.1093/jhered/93.1.77

Ward, J.A., J. Bhangoo, F. Fernández-Fernández, P. Moore, J.D. Swanson, R. Viola, R. Velasco, N. Bassil, C.A. Weber, and D.J. Sargent. 2013. Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. BMC Genomics 14:2. doi:10.1186/1471-2164-14-2

Wright, L., and A. Turhollow. 2010. Switchgrass selection as a "model" bioenergy crop: A history of the process. Biomass Bioenergy 34:851–868. doi:10.1016/j.biombioe.2010.01.030

Wu, J., J.N. Jenkins, J.C. McCarty, and X.-Y. Lou. 2011. Comparisons of four approximation algorithms for large-scale linkage map construction. Theor. Appl. Genet. 123:649–655. doi:10.1007/s00122-011-1614-8

Wu, J., T. Maehara, T. Shimokawa, S. Yamamoto, C. Harada, Y. Takazaki, N. Ono, Y. Mukai, K. Koike, J. Yazaki, F. Fujii, A. Shomura, T. Ando, I. Kono, K. Waki, K. Yamamoto, M. Yano, T. Matsumoto, and T. Sasaki. 2002. A comprehensive rice transcript map containing 6591 expressed sequence tag sites. Plant Cell 14:525–535. doi:10.1105/tpc.010274

Wu, K.K., W. Burnquist, M.E. Sorrells, T.L. Tew, P.H. Moore, and S.D. Tanksley. 1992. The detection and estimation of linkage in polyploids using single-dose restriction fragments. Theor. Appl. Genet. 83:294–300. doi:10.1007/BF00224274

Yang, J., B. Benyamin, B.P. McEvoy, S. Gordon, A.K. Henders, D.R. Nyholt, P.A. Madden, A.C. Heath, N.G. Martin, G.W. Montgomery, M.E. Goddard, and P.M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42:565–569. doi:10.1038/ng.608

Young, H.A., B.J. Hernlem, A.L. Anderton, C.L. Lanzatella, and C.M. Tobias. 2010. Dihaploid stocks of switchgrass isolated by a screening approach. BioEnergy Res. 3:305–313. doi:10.1007/s12155-010-9081-7

Young, H.A., G. Sarath, and C.M. Tobias. 2012. Karyotype variation is indicative of subgenomic and ecotypic differentiation in switchgrass. BMC Plant Biol. 12:117. doi:10.1186/1471-2229-12-117

Yu, J.K., T.M. Dake, S. Singh, D. Benscher, W. Li, B. Gill, and M.E. Sorrells. 2004. Development and mapping of EST-derived simple sequence repeat markers for hexaploid wheat. Genome 47:805–818. doi:10.1139/g04-057

Yuan, J.S., K.H. Tiller, H. Al-Ahmad, N.R. Stewart, and C.N. Stewart. 2008. Plants to power: Bioenergy to fuel the future. Trends Plant Sci. 13:421–429. doi:10.1016/j.tplants.2008.06.001

Zamir, D., and Y. Tadmor. 1986. Unequal segregation of nuclear genes in plants. Bot. Gaz. 147:355. doi:10.1086/337602