PROJECT FINAL REPORT "DE-SC0008781"

DECIPHERING NATURAL ALLELIC VARIATION IN SWITCHGRASS FOR BIOMASS YIELD AND QUALITY USING A NESTED ASSOCIATION MAPPING POPULATION

Participating Organizations

The Samuel Roberts Noble Foundation, Inc.
2510 Sam Noble Parkway, Ardmore, OK 73401

University of Tennessee
2431 Joe J. Dr., Knoxville TN 37996

University of Wisconsin
1575 Linden Dr., Madison, WI 53706

USDA-ARS, Dairy Forage Research Center
1925 Linden Dr., Madison, WI 53706

National Renewable Energy Laboratory
1617 Cole Blvd., Golden, CO 80401

Principal Investigator

Malay C. Saha, Associate Professor, Forage Improvement Division (FID)
The Samuel Roberts Noble Foundation, Inc. 2510 Sam Noble Parkway, Ardmore, OK 73401
Phone: 580-224-6840, Fax: 580-224-6802, Email: mcsaha@noble.org

Co-Investigators

E. Charles Brummer, Director and Senior Vice President, FID, Noble Foundation
Shawn Kaeppler, Professor, University of Wisconsin
Hem S. Bhandari, Assistant Professor, University of Tennessee

Abstract: Switchgrass (*Panicum virgatum* L.) is a C4 grass with high biomass yield potential and a model species for bioenergy feedstock development. Understanding the genetic basis of quantitative traits is essential to facilitate genome-enabled breeding programs. The nested association mapping (NAM) analysis combines the best features of both bi-parental and association analyses and can provide high power and high resolution in QTL detection and will ensure significant improvements in biomass yield and quality. To develop a NAM population of switchgrass, 15 highly diverse genotypes with specific characteristics were selected from a diversity panel and crossed to a recurrent parent, AP13, a genotype selected for whole genome sequencing and parent of a mapping population. Ten genotypes from each of the 15 F1 families were then chain crossed. Progenies form each family were randomly selected to develop the NAM population. The switchgrass NAM population consists of a total of 2000 genotypes from 15 families. All the progenies, founder parents, F1 parents (n=2350) were evaluated in replicated field trials at Ardmore, OK and Knoxville, TN. Phenotypic data on plant height, tillering ability, regrowth, flowering time, and biomass yield were collected. Dried biomass samples were also analyzed using prediction equations of NIRS at the Noble Foundation and for lignin content, S/G ratio, and sugar release characteristics at the NREL. Genomic shotgun sequencing of 15 switchgrass NAM founder parental genomes at JGI produced 28-66 Gb high-quality sequence data. Alignment of these sequences with the reference genome, AP13 (v3.0), revealed that up to 99% of the genomic sequences mapped to the reference genome. A total of 2,149 individuals from NAM populations were sequenced by exome capture and two sets of 15 SNP matrices (one for each family) were generated. QTL associated with important traits have been identified and verified in breeding populations. The QTL detected and their associated markers can be used in molecular breeding programs to facilitate development of improved switchgrass cultivars for biofuel production.

Project objectives

The goal of the project is to understand the genetic basis of the key biofeedstock traits of biomass yield and biomass composition in order to accelerate development of superior cultivars. Switchgrass is a perennial species requiring multiple years to complete a breeding cycle. Molecular markers closely associated with quantitative trait loci (QTL) for major traits could enable marker-only selection based on genotyping seedlings in the greenhouse, potentially improving gain. In order to realize this potential, we proposed three specific objectives:

> ***Objective 1: Develop a nested association mapping (NAM) population of 2,000 plants and construct a genetic map for this population.*** Effective marker-trait associations require that markers are close to, or actually within, the gene affecting phenotype. A NAM population provides more refined QTL localization and a broader genetic diversity than commonly used biparental populations.
>
> ***Objective 2: Identify QTLs and molecular markers associated with biomass yield, feedstock quality and other agronomically important traits.*** The critical biofuel traits of biomass yield and composition was evaluated in field trials at two diverse locations Phenotypes were associated with genetic markers to identify QTL.
>
> ***Objective 3: Validate marker-QTL associations in breeding populations.*** Markers identified in Objective 2 were verified in breeding populations developed in our program to determine their potential for marker-assisted selection (MAS).

Project outcomes

Development of a nested association mapping population

The nested association mapping (NAM) analysis combines the best features of both bi-parental and association analyses and can provide high power and high resolution in QTL detection and will ensure significant improvements in biomass yield and quality. A NAM population consists of interlinked biparental crossing populations that collectively sample a broader set of germplasm than a biparental population but does not require higher number of markers necessary for association mapping. Fifteen diverse switchgrass genotypes with distinct characteristics were selected from a diverse collection of 360 genotypes from 36 accessions consists of GRIN collection and breeding populations. The selected genotypes were crossed to the recurrent parent AP13, a genotype selected for whole genome sequencing and parent of a mapping population. Chain cross schemes were followed with 10 F1 individuals selected from each of the 15 F_1 families (Fig. 1). Parent specific SSR markers were developed and the hybrids were confirmed using these markers. Seedlings from each of the chain cross (pseudo F_2 progenies) were grown in the Noble Foundation greenhouse and a population of 2,000 plants were randomly selected (200 progenies from seven families and 75 from eight families) to construct the switchgrass NAM population of 2000 progenies.

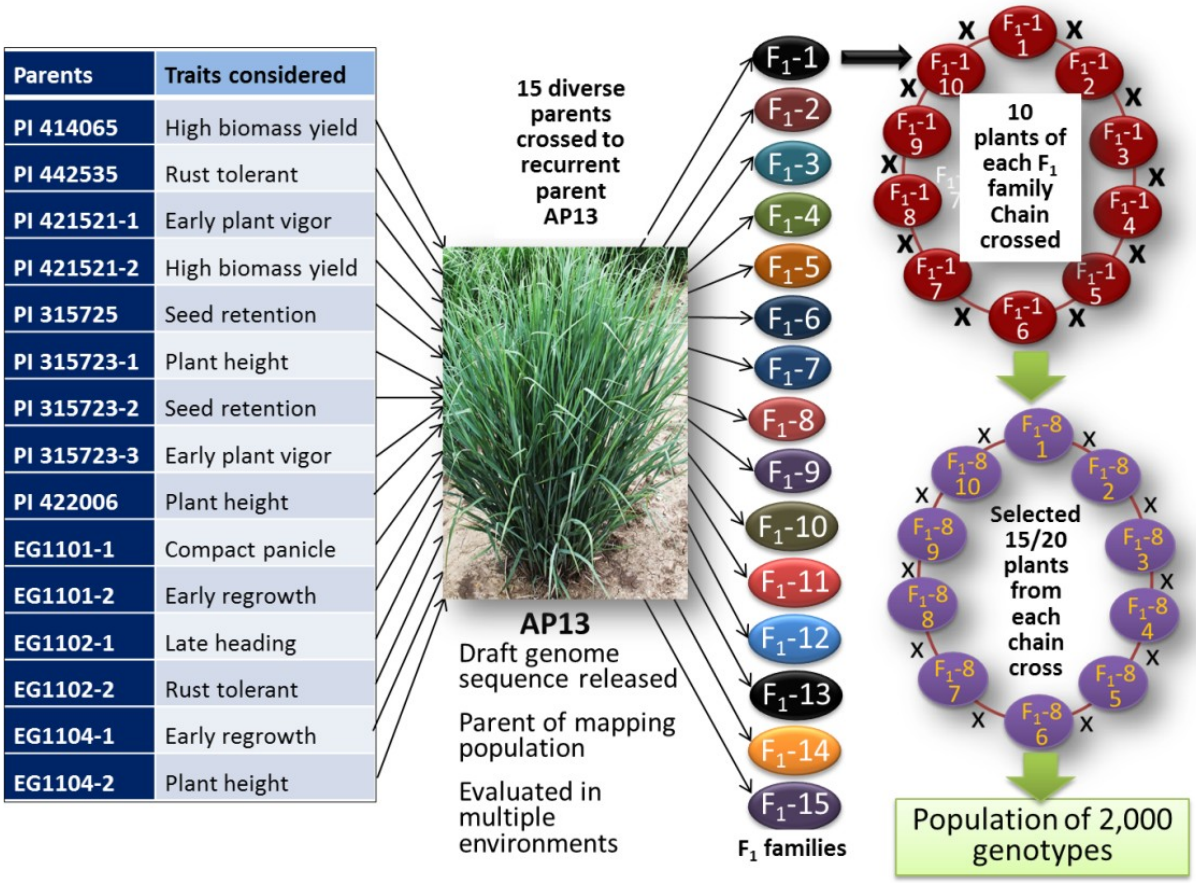

Figure 1: Crossing scheme for the development of a switchgrass NAM population. A total of 15 diverse parents selected from natural populations were crossed to the recurrent parent, AP13. The population consists of 2,000 pseudo F_2 progeny.

Field evaluation and phenotypic data collection

All the founder parents, F1 parents, progenies were multiplied into clonal ramets which were then planted in fields of Ardmore, Oklahoma and Knoxville, Tennessee (Fig. 2). In 2013, we were able to produce enough ramets for planting one replication in each location. Two other replications were planted in 2014. Field planting was conducted following the Alpha Lattice design. A total of 2,350 plants, as outlined below, were accommodated in each replication.

1.	200 genotypes from each of seven pseudo F_2 families:	1,400
2.	75 genotypes from each of eight pseudo F_2 families:	600
3.	Multiple copies of AP13	30
4.	Three copies of each of 15 grandparents	45
5.	Two copies of each of chain cross F_1 parents (135 x 2)	270
6.	Five copies of an 'Alamo' check	5
	Total	2,350

Figure 2: NAM population evaluation at Knoxville, TN (A) and Ardmore, OK (B).

The population was established well at both locations and was harvested after senescence at each year. Phenotypic data were collected on tiller density, panicle color, plant height and biomass yield. In addition, stem rust reactions were scored at Knoxville. The plants in Ardmore didn't show reasonable stem rust symptoms. Wide variability was observed in the population for many traits (Fig. 3).

Figure 3: Distinct phenotypic and family variations have been observed in the NAM population for many traits.

Significant variations were observed for biomass production within and among families at both locations (Fig. 4). Distinct heterotic effects (up to 56% above parent) have been observed in several families (Fig. 5). Total shoot biomass including stem, leaf and panicle harvested after senescence, was sampled from both locations in 2014 and 2015. Samples were oven dried at approximately 40ºC for 72 hours and milled to a 1 mm particle size (mesh size 20) using a Thomas Wiley® Mill Model 4 (Thomas Scientific, Swedesboro, NJ, USA). The milled samples were analyzed for lignin content and sugar (glucose and xylose) release at the National Renewable Energy Laboratory (NREL), Golden, CO.

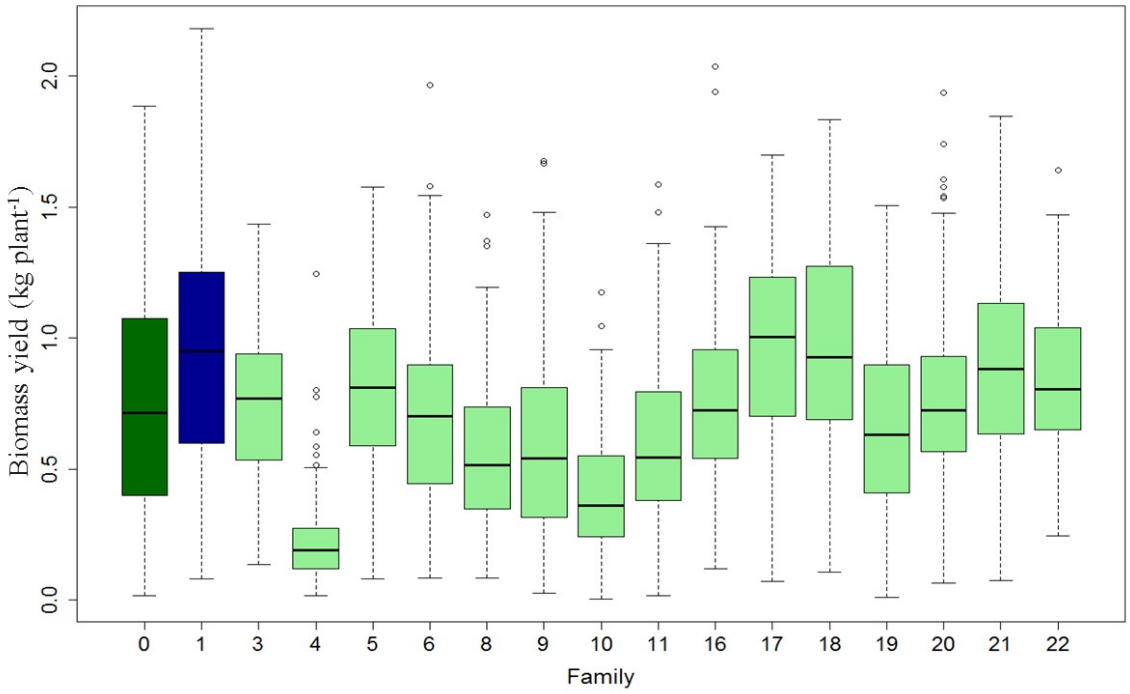

Figure 4: Biomass distribution of the NAM population. Data collected from the Ardmore field in 2014 growing season.

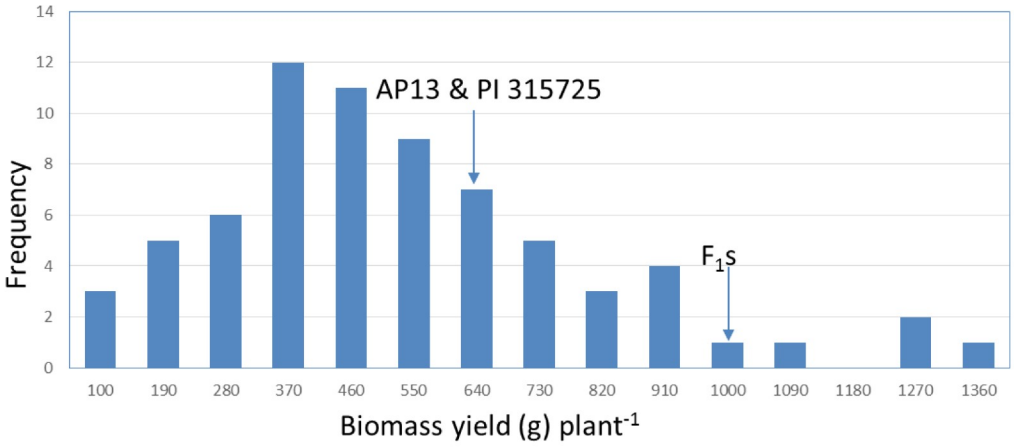

Figure 5: High heterosis (56% above parents) was observed in AP13 x PI315725 crosses, when evaluated in Ardmore, OK field.

Genotyping and mapping of the founder parents

Genomic shotgun sequencing of 15 switchgrass NAM parental genomes produced 28-66 Gb high-quality sequence data. Alignment of these sequences with the reference genome, AP13 (v3.0), revealed that up to 99% of the genomic sequences mapped to the reference genome. The parent, NFGA16_05, produced the highest number (9.94 million) of polymorphic SNPs whereas, the least polymorphic SNPs (6.43 million) were observed in NFGA09_02. We cataloged 27.78 million bi-allelic SNPs in the 18 chromosomes of a tetraploid switchgrass genome. On an average one SNP was identified in every 48 to 64 bp of chromosome sequence of the NAM parental genomes with regional fluctuations were observed in the SNP density within each chromosome (Fig. 6). The ration of intronic to exonic SNPs was 1.72. We have identified 1.09 million nonsynonymous SNPs in the exonic regions of NAM parental genomes with 7,128 SNP dense genes.

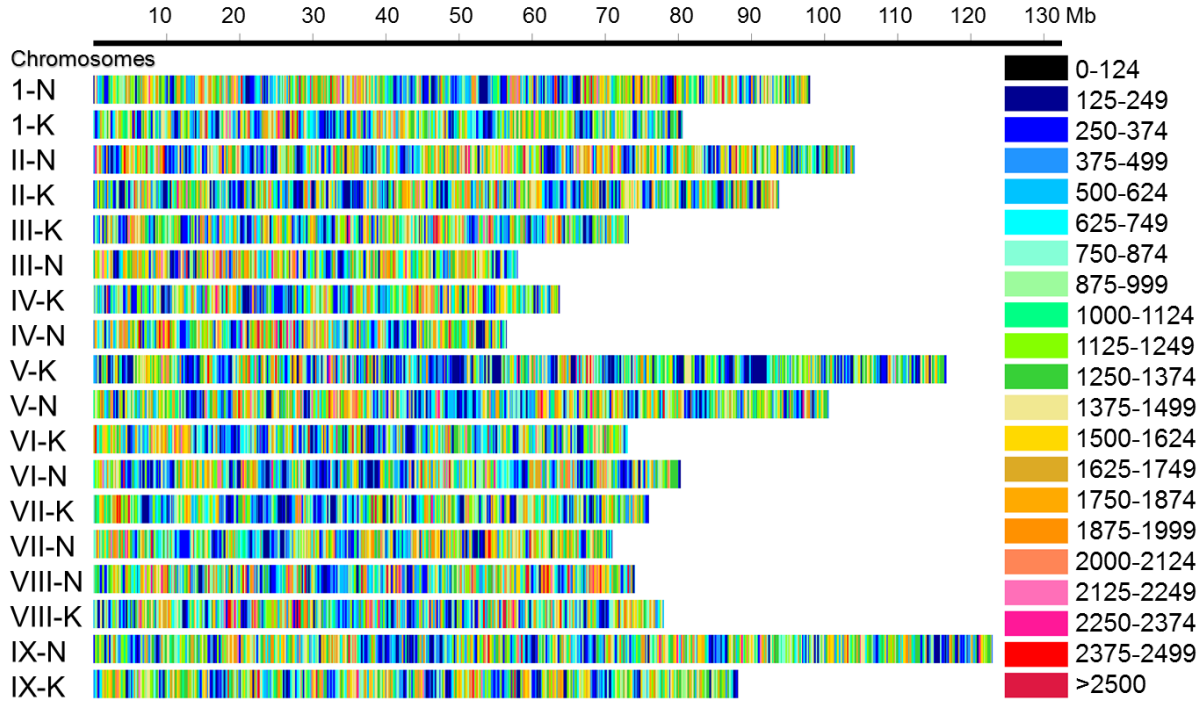

Figure 6: Color coded SNP frequency distribution in every 50 Kb interval across 18 chromosomes.

Genotyping and construction of genetic maps in NAM population

The NAM population was genotyped following the exome capture protocol. Genetic linkage maps were constructed and used in trait mapping.

Task 1: Sequence exome capture features from each individual in the entire NAM pedigree, for single nucleotide polymorphism (SNP) detection.

Task 1 Accomplishments: A total of 2,149 individuals were sequenced by exome capture. This includes: 5 AP13 samples, the 15 diverse founders, 128 F1 hybrid parents, and 2001 F2 progeny.

Task 2: Map sequences to switchgrass reference genome (v.1.1) and identify SNPs.
Task 2 Accomplishments: Two sets of 15 SNP matrices (one for each subpopulation) were generated:

1) "Hapmap" set, based on a large, diverse panel of switchgrass accessions
2) "Custom" set, based solely on the NAM population

Each "hapmap" SNP matrix contains over 1.2 million SNPS and each "custom" matrix has over 540,000 SNPs.

Task 3: Filter SNPs and call the genotypes for each NAM population individual at each SNP locus.
Task 3 Accomplishments: Used "custom" SNP matrices for genotype calling. Filtered SNPs based on % missing data, polymorphism within each subpopulation, segregation ratio, and read depth. Created a set of "true" segregating SNPs for each of eighteen chromosomes, with an average of 1,900 SNPs per subpopulation.

Task 4: Group SNPs into linkage groups and calculate genetic distances for map construction
Task 4 Accomplishments: JoinMap (v4.1) software was employed to group a set of 1,500 – 3,500 SNPs per subpopulation into linkage groups. Groups that were created using recombination frequency matched the physical chromosome assignments of the included SNPs, based on mapping (Task 2). Linkage groups from each subpopulation were merged with respective groups, based on SNP chromosome assignment, and the regression-mapping algorithm was used to calculate the order of and genetic distance between the SNPs, for map construction.

Task 5: Genetic Linkage Map Construction
Task 5 Accomplishments: Genetic linkage maps were constructed. Initial framework map consists of 18 linkage groups with an average of 200 SNPs per group (Fig. 7). Final linkage map was constructed with 2,684 SNP markers.

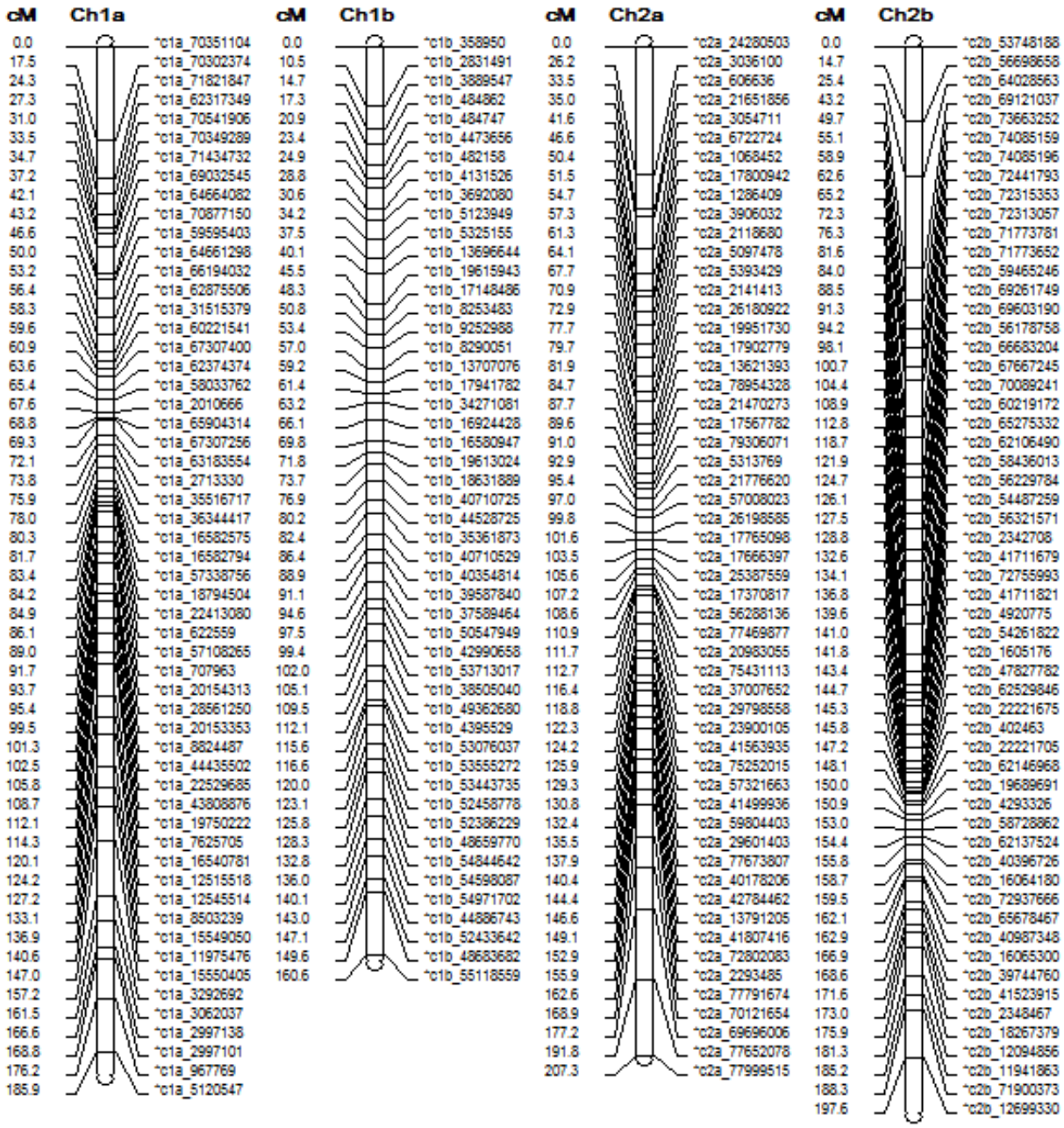

Figure 7. Genetic linkage maps of chromosomes 1a, 1b, 2a and 2b of switchgrass NAM population composed of average 200 SNP markers.

QTL mapping for the biomass yield was performed using WinQTLCartographer. The NAM population map consists of 2,684 SNP markers and biomass yield data (g) for all 2,000 pseudo-F2 genotypes across different environments were used for the QTL detection. Data obtained from each year within a location and across locations were analyzed. Composite Interval Mapping using regression procedure resulted 53 QTLs which have positive additive allelic effects. Biomass QTL were identified in most of the chromosomes except I-a, II-b, III-a

and VII-b. Thirteen QTLs repeated in either multiple years and/or locations are presented in Table 1. Some of the QTLs appeared in the same chromosomal position in multiple years and locations are presented in Fig. 8.

Table 1: Biomass QTLs identified across years and/or locations and their distributions in switchgrass chromosomes.

Common Biomass QTL	I-b	IV-a	IV-b	V-a	VI-b	VII-b	VII-a	IX-a	IX-b
Ard2013 & Ard2014							1		
Ard2013 & Ard2015	1		2						
Ard2014 & Ard2015				1					
Ard2013, Ard2014 & Ard2015		1		1	1				1
Ard2013 & Kn2015									1
Ard2014 & Kn2014								1	
Ard2015 & Kn2015						1			
Ard 2014, Ard2015 & Kn2014									1

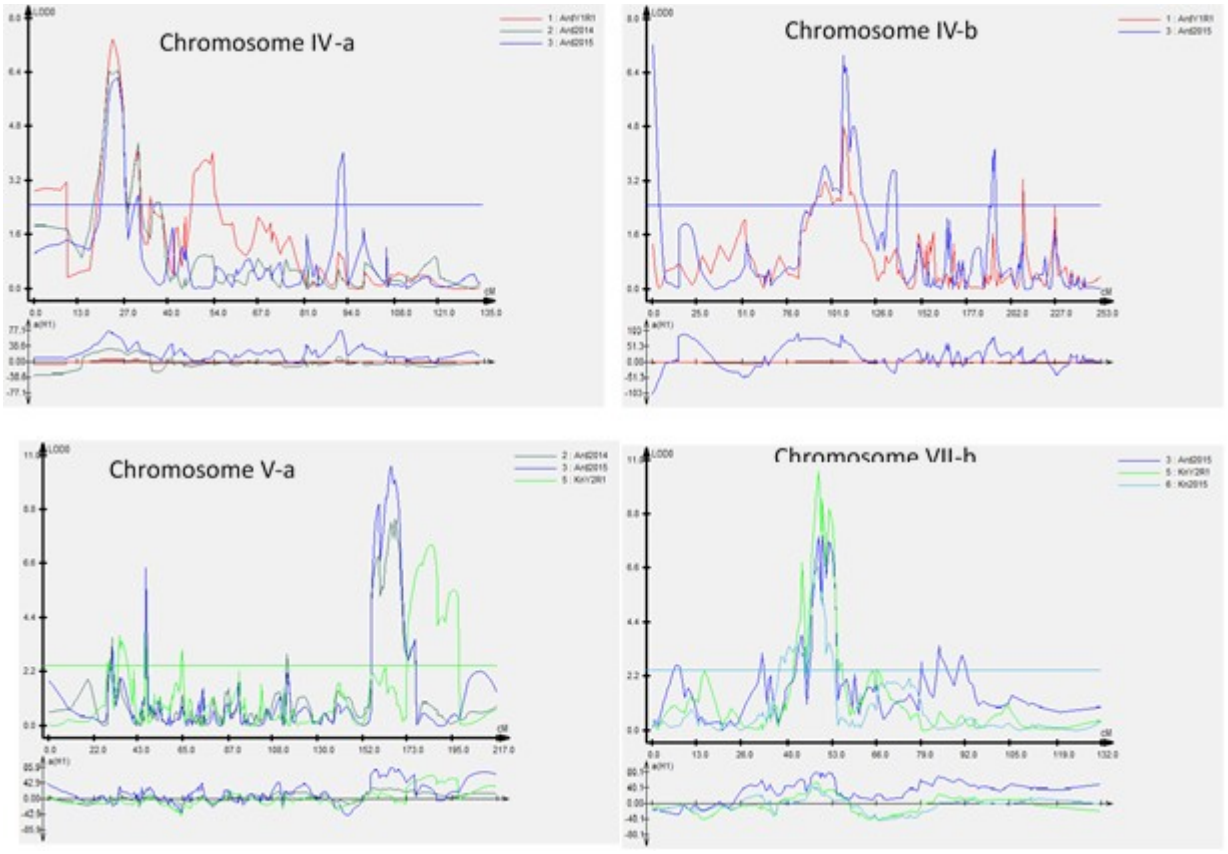

Figure 8: QTLs in the same chromosomal location identified in multiple years and locations.

Validation of biomass QTL

Markers associated with three important biomass QTL were verified in a switchgrass association mapping population. Genotypes with high biomass potential carries all the alleles associated with high biomass.