Development of Genomic and Genetic Tools for Foxtail Millet, and Use of These Tools in the Improvement of Biomass Production for Bioenergy Crops

Project ID 0014514
Prog Mgr Catherine M. Ronning
PI: Andrew Doust
Award Register#: DE-FG02-08ER64636

Aims and Objectives

The overall aim of this grant was to develop genomic and genetic tools in foxtail millet that will be useful in improving biomass production in bioenergy crops such as switchgrass, napier grass, and pearl millet. A variety of approaches have been implemented, and our lab has been primarily involved in genome analysis and quantitative genetic analysis. Our progress in these activities has been substantially helped by the genomic sequence of foxtail millet produced by the Joint Genome Institute (Bennetzen et al., in prep). In particular, the annotation and analysis of candidate genes for architecture, biomass production and flowering has led to new insights into the control of branching and flowering time, and has shown how closely related flowering time is to vegetative architectural development. These are of direct significance to bioenergy crops such as switchgrass, where it is important to maximize vegetative growth for greatest biomass production.

Specific objectives pursued in this grant included:

1. Conversion of genes and RFLP markers into SNP-based markers for mapping and association analysis.
2. Genotyping of an F_7 recombinant inbred line (RIL) population; using SSR and SNP based markers.
3. QTL analyses of the RIL population, including:
   i. Analysis of architecture, biomass and flowering time in greenhouse and field experiments.
   ii. Developmental QTL analysis of branching and other biomass traits in the greenhouse and in the field.
   iii. Analysis of the effect of crowding on branching and other biomass traits in field grown populations.
4. Developmental analyses of foxtail and green millet to investigate branching potential under low and high planting densities.
   i. Comparative genomics, including annotation of candidate plant architecture, flowering time, and biomass genes, and phylogenetic analysis of candidate gene evolution across the grasses
5. Experimental crosses between foxtail and green millet – not discussed here as this objective was achieved by co-PIs Devos and Bennetzen.

Marker design, genotyping, and map creation for QTL analysis

We constructed a genetic map of an F7 RIL population derived from a cross between foxtail (*Setaria italica*) and green (*S. viridis*) millet, developed by Katrien Devos (co-PI and Olivier Panaud. This used 102 SSRs (using published primers), 561 SNPs distributed throughout genome (from SNP genotyping conducted in PI Devos lab), 17 SNPs designed from RFLPs used in the first published foxtail millet genetic map (Wang et al., 1998, Theoretical and Applied Genetics), and 6 SNPs designed to specific candidate genes for biomass traits. The SNP data was donated by co-PI Devos, while all other markers were used by us to genotype the RIL population. All markers have a unique position in the map, and clustered markers were removed. All markers are anchored to the genome sequence, allowing previous genetic maps constructed with SSRs and RFLPs to be readily compared to this map. This map is our primary tool for quantitative trait locus (QTL) analysis (Fig. 1).

Quantitative genetic analysis

Our quantitative genetic research has focused on plant architecture, biomass, and flowering time. We have measured multiple traits at various times during development of the plants in an F7 RIL population kindly provided by Dr Olivier Panaud and Dr Katrien Devos (co-PI on grant) (Table 1). Measured traits included: tiller number, aerial branch number, height of culm and tillers, node number on culm, orders of vegetative branching, days to flowering, and biomass at harvest (dry weight).

Table 1. Growth trials of F7 RIL population at Oklahoma State University, Stillwater, OK

Date of trial	Environment	Developmental stages measured
Summer 2008	Greenhouse	Seven stages throughout development, including two weeks after planting, first flowering and harvest
Late Spring 2010	Field	Two weeks after planting, first flowering and harvest
Summer 2010	Greenhouse: Low and high density plantings	Two weeks after planting, first flowering and harvest
Late Spring 2011	Field: Low and high density plantings	Two weeks after planting, first flowering and harvest

QTL analyses of architecture and flowering time: Individual QTLs found for differences in the biomass-related traits of branch number and height in a greenhouse trial conducted in 2008 and one conducted in 2010 are as follows (Fig. 1):

1. 2008
   - Height: 8 QTLs explaining 65% of the phenotypic variance. One on Linkage Group (LG) 9 explains 13.8% and one on LG5 explains 9.1%.
   - Tiller number: 5 QTLs explaining 47.4 % of the phenotypic variance.
   - Days-to-flowering: 7 QTLs explaining 88.6 % of the phenotypic variance. Two on LG5 explain 16% and 16.5% of the phenotypic variance respectively (they are not the same QTL).
2. 2010

- Height: 6 QTLs explaining 50.7% of the phenotypic variance. One on LG9 explains 13.7% and one on LG5 explains 8.2% of the phenotypic variance.
- Tiller number: 7 QTLs explaining 56.1 % of the phenotypic variance.
- Days-to-flowering: 5 QTLs explaining 60.8 % of the phenotypic variance. Two on LG5 explain 10.9% and 18% of the phenotypic variance correspondingly (they are not the same QTL).

The amount of phenotypic variance explained by these QTL vary between trials, so that, for example, in 2008 one on LG5 explains 21% of the phenotypic variance but in 2010 it accounts only for 7%.

QTL analyses indicate that all of the QTL associated with flowering time are also associated with QTL for either branching or height, although the reverse is not true (not all QTL for architectural traits are necessarily associated with flowering time) (Fig. 1). These results are the focus of our first QTL publication from this grant, planned for submittal later this year. A second focus of the QTL work has been on flowering time across multiple environments and photoperiod regimes (particularly short day versus long day growing conditions), and has revealed that flowering appears to be repressed in long day conditions. A major QTL for flowering on chromosome 4, whose underlying candidate gene appears to be Flowering Locus T (FT), is revealed in short day conditions but absent in long day conditions.

Developmental QTL analyses: In a greenhouse trial in 2008, architectural trait measurements were taken at several time periods both before and after flowering. QTL analyses showed that QTL peaks could shift between time periods, implying that the same phenotypic trait could be under the control of multiple and shifting gene sets (Fig. 2). For example, panel A and B show that the same QTL for tillering can have different levels of significance at different stages of development, while panels C and D show that different QTL are found between early developmental stages and at flowering and harvest for height.

Density trials: Analysis of the two trials (Greenhouse 2010 and Field 2011) that examined effect of density on architecture and biomass are ongoing, with all of the phenotyping being completed and preliminary QTL analyses completed. These analyses show that QTL effects can be similar between densities, but more pronounced in the low density plantings, although there are some QTL which are only found in the low density plantings (Fig. 3).

Analysis of differences in biomass accumulation between different density plantings reveals that most QTL are the same, although the relative importance of the QTL varies significantly between densities (Fig. 4). For example, the highly significant QTL on chromosome 9 associated with height and branching genes such as dwarf8 and teosinte branched1 in the high density trial is hardly significant in the low density trial. Similarly some QTL for vegetative biomass map to the same regions as those for inflorescence biomass, whereas in other cases QTL for these two traits map to other regions. Comparing the QTL for biomass with those for branching and height (Figs. 1 and 4)

shows that these are highly correlated, and make obvious targets for breeding in foxtail millet and, by extension, bioenergy grasses such as switchgrass.

Figure 1. Genetic map of Setaria, with QTL for architectural and flowering time traits in a greenhouse and a field trial. Bars represent QTL ± 95% confidence interval, established through analysis of 1000 permutations of the data set.

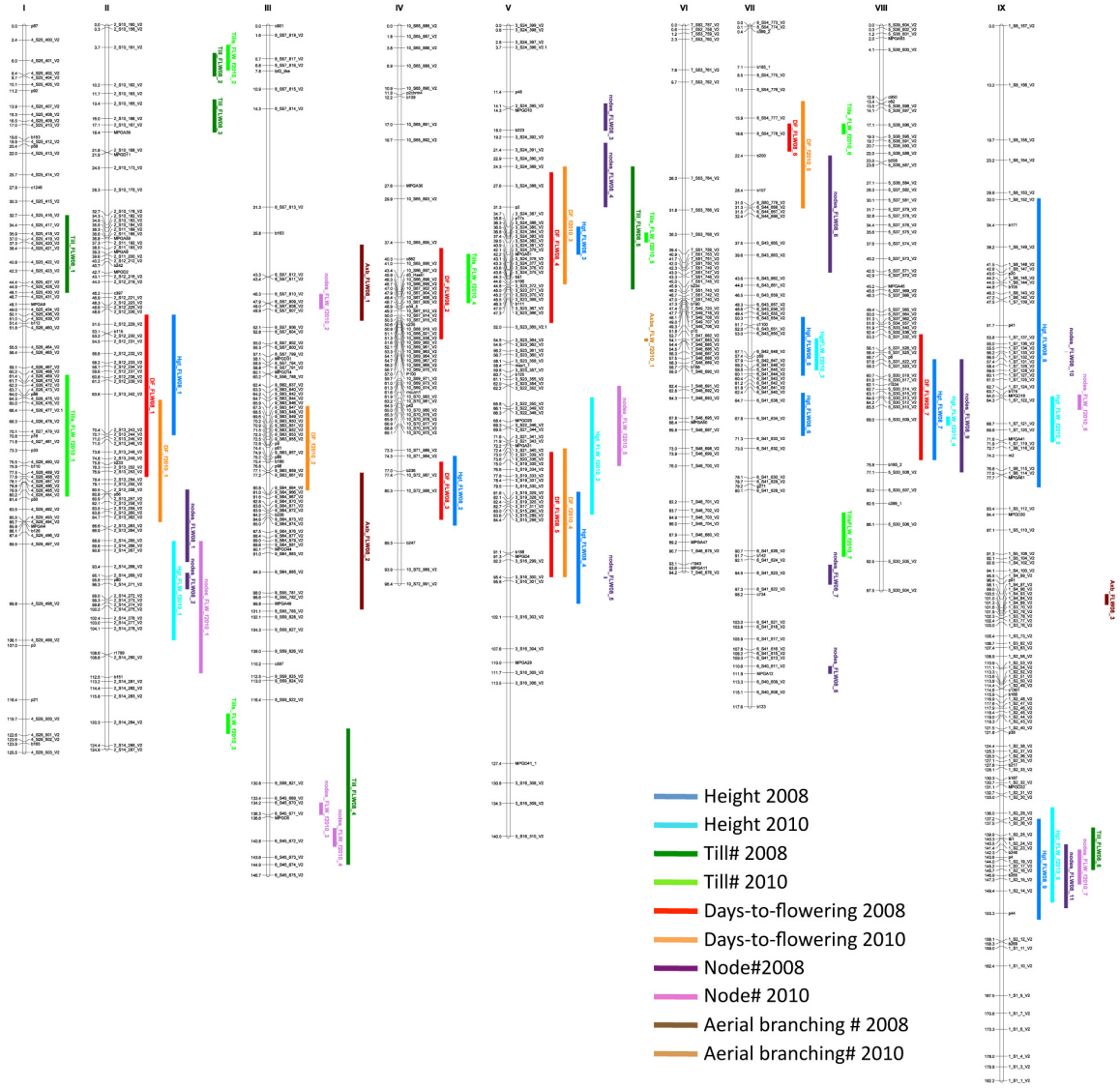


Figure 2. Developmental QTL analyses showing that QTL for the same trait differ between time periods. A. Top portion of graph shows tiller number QTL for all nine chromosomes (laid end to end), peaks above horizontal bar are significant (calculated by 1000 permutations of the data). Bottom portion shows size and direction of effect for each QTL. FLT= flowering, HVT = harvest, T1-T5 are five measuring points between germination and harvest. B. Tiller number QTL on chromosome 9, showing that the same QTL position may be more or less significant depending on the sampled time period. C. Height QTL for all nine chromosomes. D. Height QTL on chromosomes 6 and 7 showing that QTL early in development (T1-T3) are not in the same position as QTL at flowering and harvest.

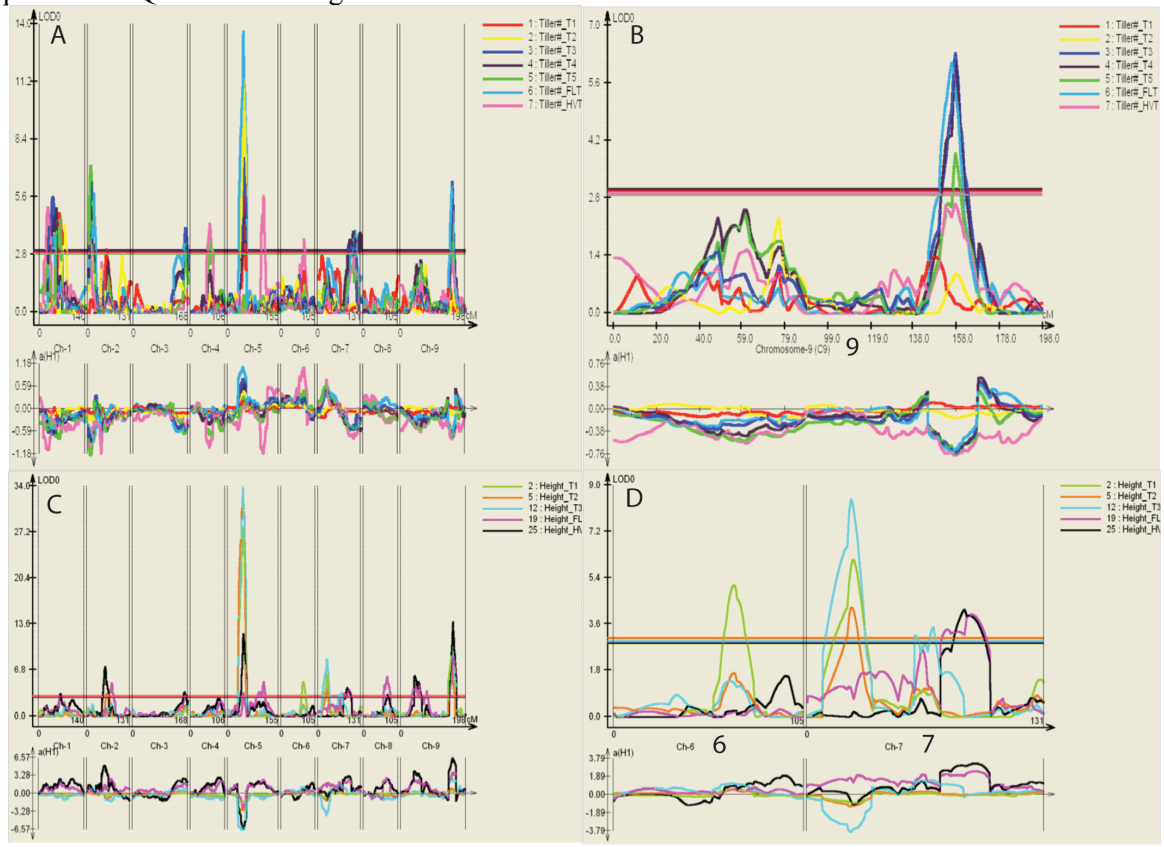


Figure 3. Flowering time, branching, and height traits in high density (upper graph) and low density (lower graph) in greenhouse trial. The X-axis represents the nine chromosomes of Setaria laid end to end, and the traits portrayed are days-to-flowering (DF), culm height (ClHgt), tiller number (Tills), aerial branch number (Axb), and days to maturity (DM, plants harvested at 50% seed maturity in the head). The horizontal black bar found on each graph corresponds to a conservative significance level, established by permutation of the data set, so that significant QTL are those peaks above the black line. The low density graph has two QTL peaks asterisked in red, these are QTL not found in the high density graph. Each graph of QTL peaks has a corresponding graph of size and direction of effect.

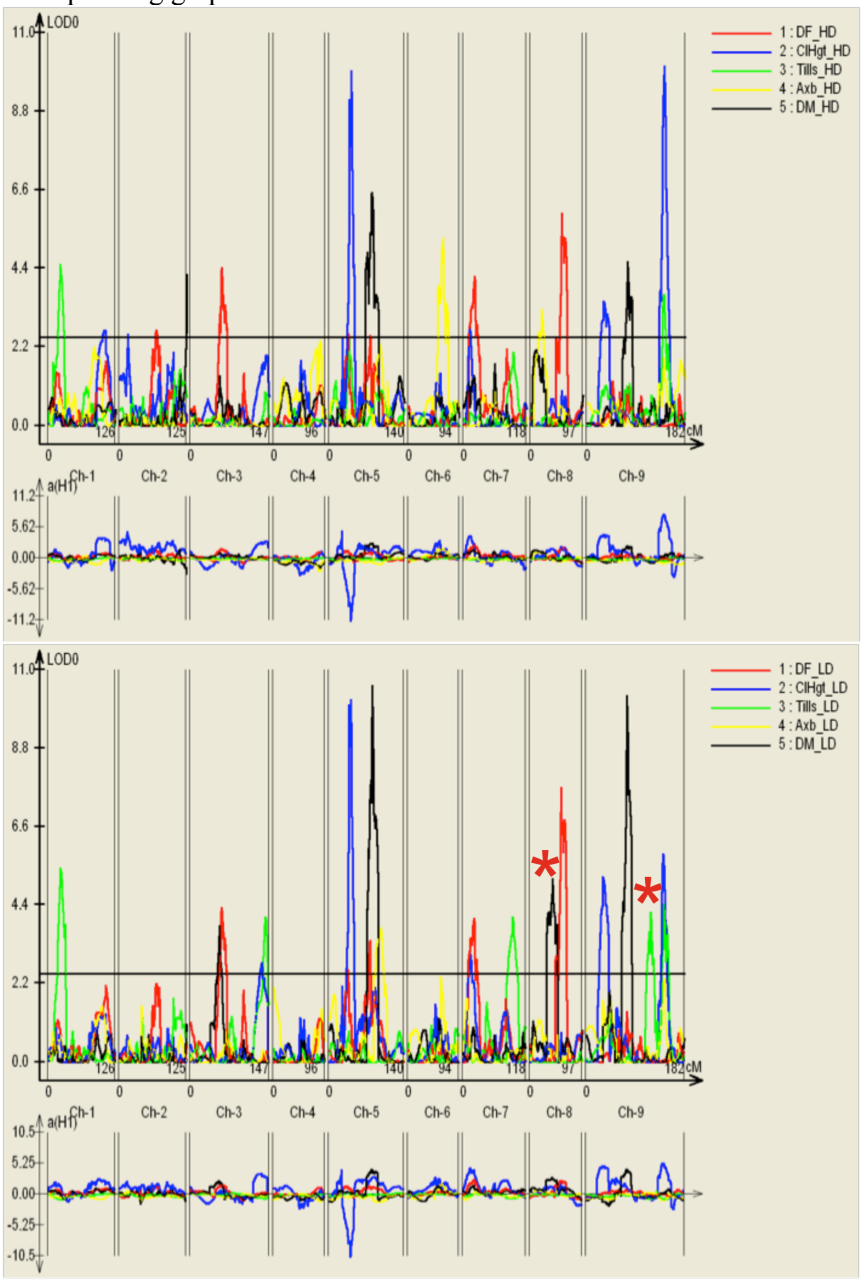
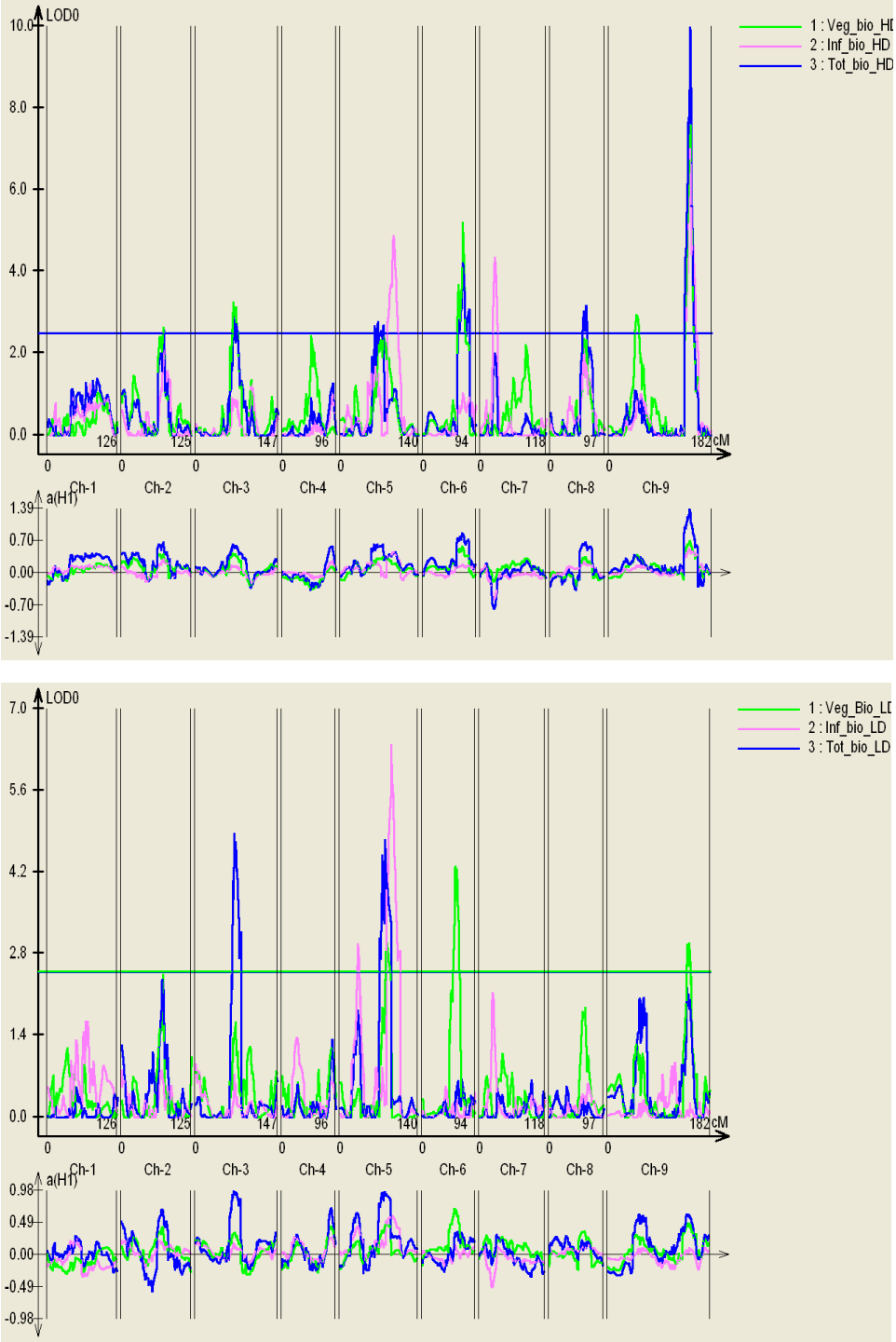


Figure 4. Biomass traits at high density (upper graph) and low density (lower graph) in greenhouse trial. The X-axis represents the nine chromosomes of Setaria laid end to end, and the traits portrayed are vegetative biomass (Veg_bio), inflorescence biomass (Inf_bio), and total biomass (combination of vegetative and inflorescence biomass, Tot_bio). The horizontal black bar found on each graph corresponds to a conservative significance level, established by permutation of the data set, so that significant QTL are those peaks above the black line.



Developmental analyses of foxtail and green millet

Green millet exhibits significantly more phenotypic plasticity than foxtail millet, as would be expected for a comparison between a wild and domesticated species. However, green millet has a characteristic pattern of growth, where the green millet parent of the RIL population exhibits approximately 7 nodes on the main culm before transition of the vegetative apical meristem to an inflorescence meristem. Foxtail millet, on the other hand, exhibits 10-12 nodes. The amount of branching and axis extension in green millet varies widely under different planting densities, but a "classic" shade avoidance response (SAR) is not seen unless the plants are in very high planting densities. At the level of densities that we chose for high density plantings some plants developed the SAR phenotype of elongated cum and reduced side branching whilst others developed into compact bushy plants where each tiller appeared to be foraging equally for light (Fig. 5). In field settings, differences in green millet at harvest between low density and high density plantings where impressive, with many more tillers and aerial branches in the low density plants (Fig. 6).

Figure 5. Diagram of phenotypes of green millet (*S. viridis*) under low and high density plantings (LD = plants 30 cm apart, HD = plants 10 cm apart).

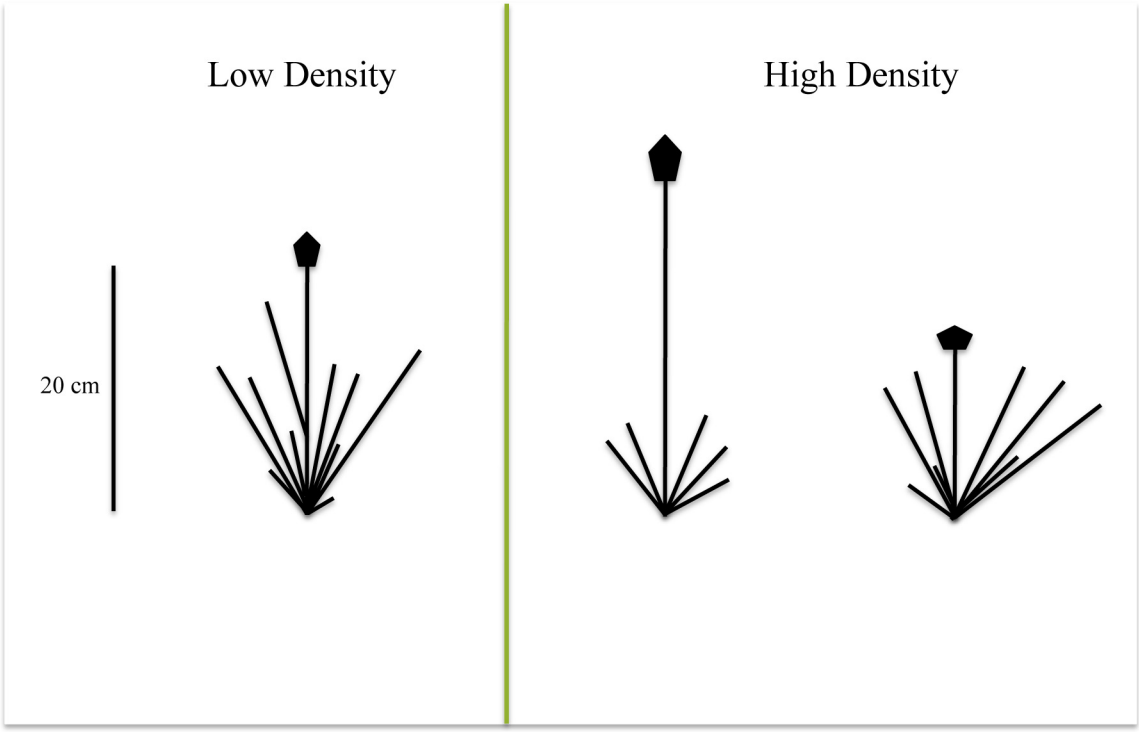
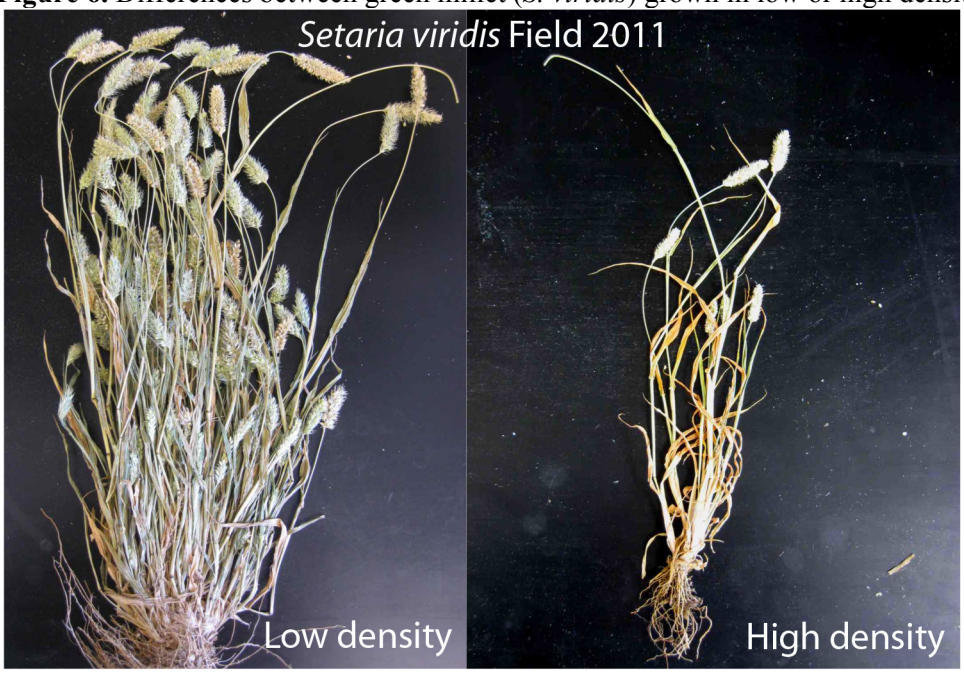


Figure 6. Differences between green millet (*S. viridis*) grown in low or high density plantings.



Comparative Genomics

The availability of the sequenced foxtail millet genome has greatly improved identification of candidate genes for the QTL, and has laid the ground-work for further intensive investigation of patterns of gene expression as well as functional analysis. Two approaches were taken, one of which was to compile lists of candidate genes from pre-existing knowledge in other species and examine if they were found in any of the QTL regions (Table 2) while the other was to investigate all genes in each QTL region and attempt to gauge if they may play a role in the affected phenotype. Homologs of candidate genes for flowering time from sequenced grass genomes (rice, Brachypodium, Setaria, Sorghum, maize) and other grasses were collected, aligned using MUSCLE and analyzed using neighbor-joining trees to identify possible Setaria orthologs for these genes (Table 3). One surprising result is that, for flowering time genes, we discovered that the photoperiod pathway in panicoids appears more similar to that of rice than to the temperate C3 grass Brachypodium. This is because the genes OsEhd1 and OsGpd7, part of a novel non-Constans photoperiod control pathway in rice, have strong orthologs in the panicoids but not Brachypodium and other pooid grasses.

Table 2. Known candidate genes for architectural and branching traits. Source: A = Arabidopsis, R = rice, T = tomato, M = maize.

Gene name	Grass species	Related genes	Growth process affected	Setaria chromosome	In QTL region?
Moc1	Rice	*LAS (A), Ls (T)*	AM initiation	4	
Bif2	Maize	*PINOID (A)*	AM initiation	9	
Ba1	Maize	*Lax1 (R)*	AM initiation	5	
Tb1	Maize	*OsTb1 (R), SbTb1 (S)*	AM outgrowth	9	Y
OsMAX1	Rice	*MAX1 (A)*	AM outgrowth	5	

D3	Rice	*MAX2 (A)*	AM outgrowth	4		
Slr1	Rice	*GAI (A), RHT (W), D8 (M)*	Plant height	9	Y	
D10	Rice	*MAX4 (A)*	AM outgrowth	5		
D14	Rice		AM outgrowth	9		
Htd1	Rice	*MAX3 (A)*	AM outgrowth	7		
Os03g0227700	Rice	*DWF4 (A)*	Height	9	Y	
Os10g0397400	Rice	*DWF1 (A)*	Height	9	Y	
Leafy head2	Rice		Height	5	Y	
OsMADS14	Rice		Height	9	Y	
OsMADS8	Rice		Height	2	Y	
Ga20Ox	Rice		Height	5	Y	
Bri1	Arabidopsis		Height	5	Y	
Gi	Rice		Height	5	Y	

Table 3. Setaria homologues for known candidate genes for flowering time (Gene list from Higgins et al. 2010 PLoS One 5:1-26). Homologues identified by reciprocal BLAST and phylogenetic analysis of grass homologs.

Gene	Arabidopsis	Rice	Brachypodium	Other cereals	Foxtail millet	Present in QTL region?
PHYA	At1g09570	Os03g51030	Bradi1g10520	HvPHYA	Si033984m	
PHYB	At2g18790	Os03g19590	Bradi1g64360		Si033968m	
PHYC	At5g35840	Os03g54084	Bradi1g08400		Si034030m	
CRY1	At4g08920	Os02g36380	Bradi3g46590		Si016503m	
CRY2	At1g04400	Os02g41550	Bradi3g49200		Si006039m	Yes, Chr. 4
CKA1	At5g67380	Os03g55389_Hd6	Bradi1g07810		Si035743m	
ZTL	At5g57360	Os06g47890	Bradi1g33610		Si006057m	Yes, Chr. 4
ELF3	At2g25930	Os01g38530			Si000443m	
	At2g25930		Bradi2g14290	TaELF3	Si021308m	
ELF4	At2g40080	Os08g27860 and Os08g27870	Bradi4g29580			
ELF4 like4 and like3		Os11g40610	Bradi4g13230			
ELF4 like4 and like3		LOC_Os03g29680	Bradi1g60090		Si038043m and	

					Si032888m	
						Yes,
GI	At1g22770	Os01g08700	Bradi2g05230	HvGI	Si000107m	Chr. 5
LHY	At1g01060	Os08g06110	Bradi3g16515		Si013398m	
CHE	At5g08330		Bradi3g60350			
TOC1	At5g61380	Os02g40510	Bradi3g48880		Si016922m	
PRR3	At5g60100	Os07g49460	Bradi1g16490	HvPpd-H1	Si033274	
COP1	At2g32950	Os02g53140	Bradi3g57670		Si016507m	
CDF1	At5g62430	Os03g07360	Bradi1g73710	HvCDF	Si035812m	
						Yes,
CO	At5g15840	Os06g16370_Hd1	Bradi1g43670	HvCO1	Si019803m	Chr. 4
SPA	At2g46340	Os01g52640	Bradi2g48660		Si000378m	
SPA	At2g46340	Os05g49590	Bradi2g15900		Si021027m	
TOE1	At2g28550	Os05g03040	Bradi2g37800	ZmRAP2.7	Si025305m	
TEM1	At1g25560	Os01g49830	Bradi2g47220		Si024576m	
Ehd1		Os10g32600			Si039992m	
ID1		Os10g28330	Bradi3g26910	ZmID1	Si036003m	
CIB1	At4g34530	LOC_Os09g29830	Bradi4g32930		Si029882m	
						Yes,
FT	At1g65480	Os06g06320_Hd3a	Bradi1g48830	HvFT1	Si008517	Chr. 4
						Yes,
HAP3A	At2g38880	Os05g38820	Bradi2g22940		Si023400m	Chr. 3
HAP5A	At3g48590	Os03g14669	Bradi1g67980		Si032636m	
GF14u	At5g16050	Os08g33370	Bradi3g36480		Si014314m	
SOC1	At2g45660	Os03g03070/100	Bradi1g77020	TaSOC1	Si039193m	
						Yes,
MADS51		Os01g69850	Bradi2g59190	TaAGL41	Si003421m	Chr. 5
LFL1		Os01g51610	Bradi2g48060		Si004459m	
LFY	At5g61850	Os04g51000	Bradi5g20340		Si011756m	

AP1	At1g69120	Os03g54160	Bradi1g08340	TaVRN1		
Ghd7		Os07g15770			Si039184m	
						Yes,
FD	At4g35900d	Os09g36910	Bradi4g36587	ZmDFL1	Si031077m	Chr. 2
		Os01g59760	Bradi2g21820	TaFDL2		
SPL	At2g33810					
FLC	At5g10140					
SVP	At2g22540	Os03g08754	Bradi1g72150	HvBM1	Si037358m	
FRI	At4g00650	Os03g63440	Bradi1g01520		Si035611m	
						Yes,
SUF4	At1g30970	Os09g38790	Bradi4g38000			Chr. 2
FCA	At4g16280	Os09g03610	Bradi4g08730	HvFCA	Si029099m	
FY	At5g13480	Os01g72220	Bradi2g60820		Si016496m	
FLD	At3g10390	Os04g47270	Bradi5g18210		Si009376m	
FVE	At2g19520	Os01g51300	Bradi2g47940		Si001403m	
FIE1	At3g20740	Os08g04270	Bradi3g14520			
MSI1	At5g58230	Os03g43890	Bradi1g13930		Si035798m	
VRN5	At3g24440	Os12g34850	Bradi4g05950	TmVIL1	Si021330m	
ARP6	At3g33520	Os01g16414	Bradi2g10130		Si017263m	
EMF2	At5g51230	Os09g13630	Bradi3g03110		Si029239m	

Publications

Publications submitted or in press

Jeffrey L. Bennetzen[1,15], Jeremy Schmutz[2,3,15], Hao Wang[1], Ryan Percifield[1,11], Jennifer Hawkins[1,12], Ana Clara Pontaroli[1,13], Matt Estep[1,4], Liang Feng[1,], Jane Grimwood[2,3], Jerry Jenkins[2,3], XXX[3], Jimmy Triplett[4, 14], Xuewen Wang[5], Xiaomei Wu[5], Xiaohan Yang[6], Chuyu Ye[6], Margarita Mauro-Herrara[7], XXX[8], Manoj Sharma[9], Rita Sharma[9], Pamela C. Ronald[9], Olivier Panaud[10], Elizabeth A. Kellogg[4], Tom Brutnell[8], Andrew Doust[7], Gerald A. Tuskan[6], Katrien M. Devos[5], and Daniel Rokhsar[3]. (submitted) Grass Genome Structure, Evolution and Adaptation Uncovered by Sequence Analysis of Setaria. Nature Biotechnology.

Jiyue Qian[1], Guanqing Jia[2], Hui Zhi[1,2], Wei Li[2], Yongfang Wang[1], Haiquan Li[1], Zhonglin Shang[3], Andrew N. Doust[4], Xianmin Diao[1,2]. (in press) Sensitivity to gibberellin of dwarf foxtail millet (*Setaria italica* L.) varieties. Crop Science

Publications in preparation

We are at present completing analyses and writing up four QTL papers. These have been delayed due to the late submission of the foxtail millet genome paper. They are:

1. a manuscript detailing a comparison between field and greenhouse grown populations for height and branching differences at flowering time.
2. A manuscript on QTL found at different stages of development and at different planting densities
3. A manuscript on the genetic control of biomass using the results of three separate field and greenhouse trials, and one on the genetics of flowering time using five separate greenhouse and field trials in three separate locations (Oklahoma, Georgia, and New York).
4. A manuscript on QTL differences in flowering time for plants grown in long and short day conditions

Conference presentations and invited seminars

Sep 2011 Doust, A.N. Parsing the grammar of grasses, vegetative branching in the millets. University of Ohio, Athens, OH.

July 2011 Mauro-Herrera, M., A.N. Doust, and M. Malahy. Vegetative architecture in grasses. International Botanical Congress, Melbourne, Australia.

Nov 2009 Doust, A.N., M. Mauro-Herrera, and M. Malahy. Branching and biomass accumulation in foxtail millet. Oklahoma Academy of Sciences, Ada, Oklahoma.

Oct 2009 Doust, A.N. Architectural evolution and domestication in grasses, University of Oklahoma, Norman, OK.

Feb 2009 Doust, A.N. Domestication and evolution of plant architecture in grasses, Kansas State University, Manhattan, KS.

Jul 2009 Mauro-Herrera, M., M. Malahy, J. Stromski, and A.N. Doust. Diversity of plant architecture in grasses: developmental stages contribute to branching patterns in foxtail millet. Botany 2009, Snowbird, Utah.

Training

Postdocs trained

Two postdocs were employed on this project, one for one year and the other for the other two years, including one Latin American and one woman. One postdoc is still with me, working on refining our understanding of the relationship between flowering, architecture and biomass production.

Graduate and undergraduate students supported

Two graduate students (masters) were supported for summer research projects, including one Native American student.
Eight undergraduate students were supported during the academic year and in the summer months, including three Native American students.

Conclusions

These analyses have identified a subset of high likelihood candidate genes for architecture, flowering time and biomass, and demonstrated the inter-relationships between these traits. The differences in QTLs identified at high and low density plantings has direct relevance to the breeding of crop grasses that are tolerant of high planting densities. The developmental analyses have shown how architecture develops and may indicate which genes may best be manipulated at various times during development to obtain required biomass characteristics. This data contributes to the overall aim of the grant to significantly improve genetic and genomic tools in foxtail millet that can be directed to improvement of bioenergy grasses.