**Working Title:** The genetic basis of two flowering time cues in switchgrass (*Panicum virgatum*)

**Authors:** Alice MacQueen\*, Li Zhang\*, Jason Bonette, …who else? People at HA who constructed the new fourway map, …, Tom Juenger

**Intended Audience:** *biologists interested in GxE, plant biologists interested in flowering*

**Target Journal:**

**Possible Friendly Reviewers/Reviewers**

**Abstract**

Switchgrass (*Panicum virgatum*) is a perennial, warm-season species native to North America which has evolved into multiple divergent populations that vary in ploidy, morphology, and phenological timing. Its potential uses have expanded in the past few decades to include biofuels. Flowering time is a key life-history trait for biofuel production, as plants cease biomass accumulation upon completion of floral development. Here, we present evidence that photoperiodic sensitivity in switchgrass flowering time varies between genetic subpopulations using a diversity panel grown at eight field sites in the central United States spanning 17 degrees of latitude. We then map the genetic basis of flowering time in this population and in an independent four-way cross created from two highly divergent southern lowland and northern upland populations. We describe eight additive QTLs across these seven field sites with moderate effects on flowering, five of which had overlapping significant associations in the diversity panel. \_sentence about major gene candidates found – prevalence of genes involved in \_\_/expressed in \_\_.

**Introduction**

The timing of floral development is a major component of plant fitness affected by multiple external environmental cues (e.g. temperature, daylength, and water availability) that signal existing or upcoming growing conditions. Genetic responses to these cues determine the speed, timing, and energy apportioned to reproductive growth, shaping both the individual’s lifespan and its lifetime production of viable seed. Global climate forcing is increasing temperatures and causing more extreme weather events, such as droughts, heatwaves, and severe rain events (Ummenhofer and Meehl, 2017). These events will alter both the timing of reproductive development and the reliability of the environmental signals that plants use to cue flowering. It is thus of increasing importance that we understand the environmental cues driving floral development and the genetics underlying flowering responses to understand the genetic potential for adaptation to novel, changing environments.

*A paragraph or two containing a summary of what is known about the physiological and molecular mechanisms of flowering in crop species. There have been good GxE flowering studies in Arabidopsis and in Helianthus – Ben Blackman’s work – as well as in annual row crops. So there aren’t many perennial systems in which GxE has been studied, and there aren’t many wild systems where we have looked at GxE in flowering, and there aren’t many outbred species where we have these studies (except Helianthus).*

*Also, an introduction to GxE and how common gardens can help study GxE. Flowering time is a canonical trait where you’d expect GxE – populations are evolving different responses to their target environments by modulating GxE in different ways. You could get GxE on the genetics side if different individuals have evolved different modulators of environmental signals. You could get GxE on the environment side if different individuals are using different cues to initiate flowering.*

Switchgrass (*Panicum virgatum*) is a warm-season perennial with wide environmental adaptation across the eastern half of North America. The US Department of Energy named switchgrass a model herbaceous biofuel feedstock in 1992

14, and since then, cultivars have been bred that significantly outproduce ethanol relative to maize and other cellulosic feedstocks15. Over the course of one season, switchgrass behaves as a determinate plant. It typically produces a single flush of tillers, which all become reproductive after a period of vegetative (leaf) development, and – critically for a biofuel crop – cease biomass accumulation upon completion of floral development (Van Esbroeck et al 2003).

*Since temperature is going to change, but light quality, photoperiod, etc. won’t, plants and breeders are going to need to figure out how to adapt to climate without botching phenology.*

Switchgrass has substantial untapped genetic and morphological diversity, with tetraploid and octoploid individuals (cite), phenotypically distinct ‘upland’ and ‘lowland’ ecotypes, and three geographically distinct, deeply diverged genetic subpopulations within tetraploid individuals (Lovell et al 20XX). Upland individuals are smaller in stature than lowland individuals, other big differences (cite). This diversity, particularly the genetic differences in flowering date and photoperiod response, can be a point of exploitation in biomass crop breeding. Breeding for plants with earlier greenup and later flowering dates may allow growers in the northern United States to take advantage of longer growing seasons, as photoperiod-sensitive strains will accumulate more biomass before flowering, contributing to higher biomass yields (cite?). Alternatively, breeding for cultivars with larger cumulative GDD requirements could increase the heat requirement needed for switchgrass development and stabilize switchgrass biomass yields at higher levels, which could help offset the negative impacts of climate warming (cite). However, despite substantial study of the mechanisms controlling switchgrass development, genetics of flowering and prediction equations with broad application remain elusive.

Here, we grow and phenotype a diversity panel of 978 distinct switchgrass genotypes, clonal replicates of which were planted at eight common garden sites across 17 degrees of latitude. We use this panel to interrogate environmental mechanisms controlling greenup and flowering. We then use a F2 cross between individuals from the most distinctive subpopulations for flowering response to dissect the genetics of flowering in these groups. Finally, we combine the results from this cross with genome-wide association results from the diversity panel to narrow in on candidate genes affecting flowering date. Taken together, our results allow us to describe the environmental cues, genes, and alleles affecting flowering across multiple distinct switchgrass populations.

**Results**

*Diversity panel captures the range of phenological responses across the common gardens*

In 2019, we grew and phenotyped a diversity panel of 978 distinct, clonally propagated switchgrass genotypes at eight common gardens. These common gardens cover the majority of the latitudinal and climatic range of switchgrass and therefore capture the most comprehensive picture to date of genotype-specific environmental plasticity, or genotype-by-environment interactions, in this species. We scored plant green up and flowering at these common gardens every two days. We then explored the overall patterns of greenup and flowering within sequenced individuals from the Midwestern subpopulation (134 genotypes) and Gulf subpopulation (229 genotypes), as these groups had the most distinct phenological responses across our common gardens (Figure 1A).

We first determined the narrow-sense heritabilities (h2)for greenup and flowering dates at single common gardens and across all eight common gardens, with common garden as a random effect (Figure 1C). To allow for the possibility that different subpopulations had different strengths of connection between these phenotypes and the genotypes (Korte and Farlow 2013), we determined h2 both within and across the Midwestern and Gulf subpopulations. Within common gardens, h2 were typically quite high: 62% on average for greenup date, and 87% for flowering date. However, h2 were not consistently high, notably for greenup at our OK and NE common gardens. Greenup dates at these sites were uncorrelated with greenup dates for clonal replicates at other sites (Figure 1B). Greenup dates at the Texas sites were strongly negatively correlated with greenup dates at the northernmost four sites. Flowering dates were generally positively correlated, and more highly correlated at the northern sites, with little correlation between TX1 and other sites in the Midwest subpopulation. These low and negative correlations were reflected in the low h2 values for models including all eight sites: h2 was 1% on average for greenup and 15.6% for flowering date. These data indicated the presence of numerous rank-changing genotype by environment interactions for these phenotypes across these common gardens.

A picture containing drawing

Description automatically generated*Figure 1. Probably need to include a map with common garden locations and locations of origin of the Midwest and Gulf individuals, maybe a timeline figure of when things happened and were measured. Perhaps this could be a strip running along the top of this current plot, with a small map on the left and a timeline on the right. Also, B and C should probably be swapped given how I talk about this data now. A) Distribution of greenup and flowering dates across two genetically distinct switchgrass subpopulations across eight common gardens. Purple represents individuals from the Midwest subpopulation, and pink individuals from the Gulf subpopulation. Vertical dashed line indicates the summer solstice. Common gardens are arranged in latitudinal order. B) Phenotypic correlations between clonal replicates planted at eight common gardens, within and between two genetic subpopulations. C) Narrow sense heritability of greenup and flowering within single common gardens (purple) and across all eight common gardens (green), within and between two genetic subpopulations*

Though we scored greenup and flowering as functions of Julian date, flowering is more likely cued by one or more temperature, rainfall, or daylength signals (*cites*). To evaluate these cues as genetic triggers of flowering, we defined greenup and flowering as functions of nine environmental cues, then determined the narrow-sense heritabilities for these phenology-related traits across our eight common garden sites. We reasoned that if a flowering trait had higher heritability than flowering as a function of Julian date, then that environmental cue was both a better predictor of flowering and was more likely to have detectable genetic variation segregating within the tested population. To allow for the possibility that different subpopulations had different cues, we also determined h2 both within and across the Midwestern and Gulf subpopulations.

Within each of the eight common gardens,

Across all eight common gardens, greenup as a function of Julian date had very low heritability within subpopulations (0.005 +/- 0.003 Gulf; 0.021 +/0 0.011 for Midwest) and across both subpopulations (h2 = 0.004 +/- 0.003). In the Gulf subpopulation, GDD 10 days before greenup had the highest heritability for greenup (h2 = 0.074 +/- 0.024), followed by the average minimum temperature for the 10 days prior to greenup (h2 = 0.065 +/- 0.022). In the Midwestern subpopulation, the average minimum temperature in the 10 days prior to greenup had the highest heritability (h2 = 0.058 +/- 0.027). Across both subpopulations, both cumulative GDD 10 days before greenup and the average minimum temperature for the 10 days prior to greenup had higher heritabilities than Julian date (h2 = 0.037 +/- 0.014; 0.036 +/- 0.014).

*<Maybe also look at just the Texas sites vs just the northern sites (without OK)>*

In contrast, flowering as a Julian date had moderate heritability within subpopulations (h2 = 0.19 +/- 0.061 Gulf; 0.097 +/- 0.045 Midwest) and moderate heritability across both subpopulations (h2 = 0.18 +/- 0.061).For the Gulf subpopulation, daylength (h2 = 0.336 +/- 0.073) was the only cue that had higher heritability than Julian date. Two environmental functions had higher heritabilities than Julian date for the Midwestern subpopulation: cumulative GDD (h2 = 0.368 +/- 0.052) and daylength (h2 = 0.144 +/- 0.052). Across both subpopulations, both daylength and cumulative GDD had higher heritabilities than Julian date (0.287 +/- 0.032; 0.201 +/- 0.041).

Because heritability for daylength varied considerably between subpopulations, we also evaluated whether plant latitude of origin affected the strength of photoperiod-related flowering signals. We observed a strong signal of latitude of origin on whether plants grown in Texas common gardens flowered in lengthening or shortening days (binomial glm, *Pr* < 2x10-16, Supp. Figure 1). The majority of plants from latitudes of origin below 35°N did not flower until days were shortening at the Texas sites, while the majority of plants from latitudes of origin above 38°N flowered while days were lengthening. Interestingly, the Oklahoma (OK) site, our first common garden at which most genotypes flowered during shortening days, was at 36°N; in contrast, the three sites where a photoperiod cue was evident were below 32°N. Plants grown in 2019 in our common gardens north of 35°N did not have sufficient vegetative growth to flower before the summer solstice, and thus were not competent to repress flowering during lengthening days. As this is a common feature of growing seasons at these latitudes, we hypothesize that plants from more northern latitudes have evolved a flowering time response to a separate, non-photoperiod based environmental cue.

If flowering date frequently varies as a function of GDD in switchgrass, this explains observations that moving southern populations northwards delays flowering, and moving northern populations south hastens flowering (Sanderson et al 1996). To explore this further, we tested whether subpopulation flowering as a function of GDD varied significantly by latitude of origin. The Midwest subpopulation response to GDD varied significantly by latitude of origin, with plants from the northernmost 20% of the range flowering at an average GDD of 568, and plants from the southernmost 20% of the range flowering at an average GDD of 779. In contrast, the southernmost and northernmost 20% of the Gulf plants differed only slightly in GDD (1008 vs 1090, \*statistically significant). Taken together, these data support a latitude-of-origin based, low GDD cue for flowering in the Midwest subpopulation, and a consistent, high GDD cue for flowering in the Gulf subpopulation that is superseded by a photoperiod cue.

*Genetic effects of greenup and flowering as functions of environmental cues*

We next evaluated genetic effects of greenup as functions of GDD and minimum temperature 10 days prior to greenup, and genetic effects of flowering as functions of daylength at flowering and GDD between greenup and flowering. To do this, we calculated BLUPs for each individual using mixed models with common garden and a kinship matrix as random effects. We then conducted genome-wide association on these BLUPs both within and between genetic subpopulations.

*Genotype-by-environment effects of flowering as functions of environmental cues*

We next evaluated the genotype by environment effects of flowering as a function of daylength at flowering and GDD between greenup and flowering. To do this, we conducted univariate GWAS for flowering at each common garden, then analyzed the allelic effects of unlinked SNPs across common garden sites for the top XK SNPs using mash, both within and between genetic subpopulations.

*Confirmation of genetic and genotype-by environment effects using a fourway cross*

The Midwest and Gulf subpopulations had the most distinct phenological responses of our five groups across our common gardens and had heritable variation for flowering in response to two distinct environmental cues. To analyze the genomic regions and allelic effects underlying the phenological responses in these subpopulations, we analyzed flowering date in an F2 cross between four individuals, two Midwest and two Gulf individuals. The parents of this cross were DAC, an early flowering Midwest individual, VS16, a late flowering Midwest individual, AP13, an early flowering Gulf individual, and WBC, a late flowering Gulf individual (Supplementary Figure X.). We made F1 crosses of the two early flowering individuals, AP13xDAC, and the two late flowering individuals, WBCxVS16. We then clonally propagated and planted the four parents, the two F1 individuals (AP13xDAC, and VS16xWBC), and 801 F2 individuals at eight field sites, then recorded greenup and flowering date every three days for the 2016-2019 seasons. Unfortunately, no parent clonal replicates flowered exclusively during shortening days at the three Texas sites (AP13: 62.4% of 173; WBC: 85.6% of 104; DAC 0% of 15; VS16 3.6% of 56).

We first compared the 2019 flowering dates of F1 individuals to the parents and the diversity panel to determine dominance of the two flowering date environmental cues. Though there was an average difference in flowering date of 17 days between the two F1 crosses, all F1 individuals flowered at similar dates as the Midwest parents each year, and as Midwest subpopulation individuals in general (Figure 2; early and late F1s flowered 8.1 +/- 8.1 and 7.9 +/- 13.3 days behind their F0 parents). In addition, neither F1 flowered consistently in shortening days in 2016 through 2019 (AxD 0% of 119; VxW 18.1% of 127). These data indicated that the lower GDD environmental cue for flowering in the Midwest subpopulation was dominant to the photoperiod response and the higher GDD environmental cue for flowering in the Gulf subpopulation.

*A close up of a map

Description automatically generated***Figure 2.** *F1 flowering dates indicate dominance of Midwest subpopulation flowering environmental cues. Vertical dashed lines show F1 flowering dates in 2019 for two Midwest x Gulf crosses. Violin plots show the distributions of flowering dates in the Midwest and Gulf subpopulations in the diversity panel. Bold yellow vertical line indicates the summer solstice. 2018 data is shown for the TX3 site, as 2019 data was not collected for this cross at this common garden.*

To determine if Gulf subpopulation flowering date phenotypes were recoverable in F2 individuals, we compared the flowering date distributions of F2 individuals to those of the parents. Very few F2 phenotypes recaptured the Gulf subpopulation parent phenotypes: only 15.7% of F2 flowering dates occurred on or after the minimum flowering date of a Gulf parent in that common garden and year. Long tails for the F2 flowering date distributions were mainly observed at TX2 and TX3, where there was the potential to isolate QTL that affected the ability of the plant to flower in response to a photoperiod cue (Figure 3, or as supplement?). F2 individuals at TX2 and TX3 had flowering dates similar to late flowering Midwest individuals to early- to mid-range flowering Gulf individuals in the diversity panel (Supplementary Figure). In contrast, at the northern five sites and at Kingsville, F2 individuals had similar flowering dates as individuals from the Midwest subpopulation. Thus, at the northern five sites, we expected to find QTL controlling flowering date from the Midwest subpopulation, potentially by affecting the cumulative GDD requirements for the plant before flowering.

**A close up of a logo

Description automatically generatedFigure 3.** *Distribution of flowering date for F2 individuals (violin plots) relative to parent plants (x’s). Maybe represent the earliest flowering date and latest flowering date for each parent as horizontal lines instead of all these X’s.*

We next analyzed the genetics of greenup and flowering in 2019 for seven fourway sites. We analyzed the flowering phenotype in two ways: as a function of Julian date (‘flowering date’), and as a function of cumulative GDD between greenup and flowering (‘flowering GDD’). There were no significant QTL for greenup. There were eight QTL for flowering date; the largest QTL were on Chr02N, Chr04K, and Chr05N (Figure 4). There were five QTL for flowering GDD, all of which overlapped with QTL for flowering date. Though most QTL LOD scores were similar for these two flowering phenotypes, the QTL on Chr02K was highly significant for flowering GDD and marginally significant for flowering date, while the QTL on the first part of Chr05N, on Chr08N and Chr09K were significant for flowering date and not significant for flowering GDD.

**Figure 4.** *Distribution of LOD scores by chromosome for flowering date and flowering GDD across 7 common gardens in 2019. Dashed lines indicate significance for the solid LOD lines of the same color.*

A picture containing colorful, white

Description automatically generatedAll eight QTL for flowering date exhibited significant GxE between common garden sites. Our cross design allowed us to estimate allelic effects of alleles from all four parents as contrasts of alleles in the F1 individuals. In the early flowering allele set, AxB, at TX2, five of eight lowland alleles delayed flowering date. In the CxD cross at TX2, four of eight lowland alleles delayed flowering date – four of the same five regions that delayed flowering date in the AxB cross. These effects were consistent with the observed order of flowering in F0 and F1 individuals. We therefore hypothesized that these regions: Chr02N, Chr04K, Chr05N, (Chr08N), and Chr09K were important regions affecting flowering in the Gulf subpopulation. At the northern sites, because of the dominance of Midwest phenotypes and alleles that we observed, we compared effects of the Midwest allele from the early and late F1 cross. For 29 site by QTL combinations for the AxB early flowering alleles, the Midwest allele was accelerating flowering by reducing flowering date. For 17 site by QTL combinations for the CxD late flowering alleles, the Midwest allele was accelerating flowering, and for 2 site by QTL the upland allele was delaying flowering. These effects, if additive, would be consistent with the observed order of flowering in the F0 and F1. All eight QTL affected flowering date for at least one northern site for both the early and late flowering allele sets. Thus, we hypothesized that all eight regions were important regions affecting flowering date in the Midwest subpopulation.

A picture containing screenshot

Description automatically generatedA screenshot of a cell phone

Description automatically generatedA picture containing hanging, light, sitting, white

Description automatically generatedAll five QTL for flowering GDD exhibited significant GxE between common garden sites. In the early flowering allele set, 27 AxB alleles accelerated flowering by decreasing the GDD required for flowering, while 20 CxD alleles accelerating flowering by decreasing GDD. No alleles delayed flowering at the northern five sites, but at TX2, four alleles accelerated flowering and two decelerated it in the AxB cross, while four accelerated it in the CxD cross. At TX1, two accelerated flowering by decreasing GDD requirements in the AxB cross, and two decelerated it, while only one allele accelerated flowering at TX1 in the CxD cross. In general, Kingsville (TX1) is a marked departure from the temperate climates of the other common gardens. Heritability for flowering was lowest at TX1 in the fourway cross (Supplementary Figure X), indicating a larger effect of environment on phenotypic variance at this site.

*Genome wide association identifies candidates affecting flowering in fourway QTL intervals*

Though QTLs detected using the fourway cross are too coarse to be informative about gene identity by themselves, combining linkage and association mapping can outperform each method used in isolation (cite Brachi, others?). We therefore conducted GWAS on greenup and flowering date at seven common garden sites for individuals from both the Midwest and Gulf subpopulation. We then analyzed the allelic effects across common garden sites for the top XK SNPs using mash. Using this data, we focused on identifying associations in or near the important QTL regions for flowering date from the fourway cross, as well as the population that allele was found within and the effect that that SNP had across sites.

Five of the eight QTL – the four largest QTL, and all five QTL that we hypothesized affected flowering in the Gulf subpopulation – colocalized with one or more of the top 25 mash hits. Discussion of candidate genes, where they are in the QTL interval, what the effects look like in mash, what subpops they are segregating in, for:

* Candidate on Chr02N
* Candidate on Chr04K
* Candidate on Chr05N 4.3 Mb
* Candidate on Chr05N 64.4 Mb
* Candidate on Chr08N

*Possible Genomic Prediction Section*

* I don’t know how to tie genomic prediction in here. Ideally we’d perhaps have information on which parts of the genome being Midwestern vs Gulf would make you sensitive to photoperiod cues, and which parts affect how much cumulative GDD you need, if you’re not sensitive to photoperiod cues. That would be a nice way to bring this full circle.

**Discussion**

* Possible reasons why we didn’t find candidates in GWAS for other three QTL intervals: false negatives, type of allele shifts we are testing here vs the fourway, other things
* And what are other types of region we can detect with GWAS that we couldn’t detect in our fourway? Particularly differences between Gulf alleles… maybe offer up Gulf specific GWAS/mash here. Or GWAS on any individual that is tetraploid and looks like it’s photoperiod sensitive – including some admixed and Atlantic individuals. As an idea.
* Want to bring back discussion of environmental cues… may be able to lead in to this by talking about which subpopulations are segregating for these things.

Spring growth is initiated by “adequate temperature” according to McMillan and Weiler (1959), with adequacy thought to be dependent on the cultivar (Parrish and Fike 2005). During vegetative growth, switchgrass phenology is closely correlated with growing degree days (GDD), the cumulative mean daily temperature less a base temperature (Madakadze et al 1998c; Sanderson and Wolf, 1995a, 1995b). A base temperature of 12 C for vegetative and reproductive development is commonly used for growth models (Kiniry et al 2005, Kinery et al 2008a, Berhman et al 2013). However, base temperatures vary by cultivar (Madakadze et al 2003) and there may be a photoperiod or vernalization mechanism rather than a temperature threshold *per se* that triggers spring growth (Parish and Fike 2005).

Switchgrass is considered a short-day plant that flowers when exposed to shortening days of a specific length (Benedict, 1940) and reproductive development is strongly linked to day-of-the year (Cornelius and Johnston, 1941; Eberhart and Newell, 1959; Hopkins et al., 1995a; Sanderson and Wolf, 1995a). However, the nature of switchgrass photoperiodicity may be genotype dependent – both northern and southern cultivars with distinctive upland and lowland ecotypes flowered under both 12 and 16 hour photoperiods, but flowering was delayed in the upland cultivar (Van Esbroeck et al 2003). Photoperiodicity likely differs with plant latitude of origin (Parish and Fike 2005). Moving plants from southern populations northward is thought to delay flowering, increasing leaf number and yields, while moving northern populations to southern latitudes is thought to hasten the transition to reproductive development, reducing vegetative growth and biomass yield (Sanderson et al 1996).

**Methods**

***Diversity panel and Single Nucleotide Polymorphism Dataset***

Overview of the diversity panel.

*Panel collection, propagation, cultivation, and phenotyping*

The formation of the diversity panel has been described previously (Lovell et al 202X). In brief, seeds, rhizomes, and clonal propagules from natural and common gardens were collected from 2015-2018, and propagated by clonal division from 2016 to 2018 with the aim of generating > 10 clones per unique accession. Plants were grown in 1 gallon pots in the final propagation before transplanting to the field. Planting at the ten field sites occurred in the spring of 2018 and followed the methods of Lowry et al 2019, with the exception that the Blackwell cultivar was used instead of the Alamo cultivar at edge positions of the plot to control for edge effects. Plant aboveground biomass was removed in the spring of 2019 before spring tiller emergence. Plants without new growth from the crown by June 1st, 2019 were removed from the experiment and replaced with Blackwell cultivar plants in July or September 2019.

<<Paragraph about how the phenotypes of greenup, emergence, and flowering were scored. Should talk with Jason for specifics after looking in his metadata about this.>>

*Panel sequencing, ploidy assessment, and ecotype classification*

The resequencing of the diversity panel has been described previously (Lovell et al 202X). Briefly, 789 diversity panel samples were resequenced at a median depth of 59x (range 20 – 140x). 630 samples were used for this analysis, after filtering for missing sequence or phenotype data, outlier heterozygousity scores, and collection site discrepancies. The reads were mapped to the V5 assembly using bwa-memcite\_bwa, and SNPs were called using SAMtools mpileup84 and Varscan V2.4.085 with a minimum coverage of eight and a minimum alternate allele count of four. Only SNPs with ≤ 20% missing data and minor allele frequencies > 0.05 were retained, resulting in 8.8 to 12.3 million SNPs used for genome wide association, depending on the individuals retained.

Ploidy assessment has been described previously (Lovell et al 202X). Briefly, two methods were used to assess ploidy: a LSRFortessa SORP Flow Cytometer (BD Biosciences), and the distribution of variant allele frequency at bi-allelic SNPs. These methods agreed for 837 out of 870 samples (96.2%) with flow cytometry data. Ecotype assessment has also been described previously (Lovell et al 202X). Briefly, at the TX2 and MI common gardens, we assessed mature switchgrass individuals at or near anthesis for a suite of 16 non-redundant traits typically used to characterize switchgrass ecotypes (e.g. leaf thickness, plant height, phenology). Convoluted neural networks were trained on seven cultivars with known ecotypes, then used to probabilistically assign ecotypes to 630 planted and sequenced tetraploid individuals. 16 tiller and leaf appearance traits were also assessed at TX2 to validate these results.

*Environmental functions for greenup and flowering*

Given our large number of sites with genetically identical individuals, we looked for an environmental cue that maximized narrow-sense heritability within or across subpopulations and common garden sites. Our reasoning was that, if a flowering time phenotype, as defined by a particular environmental cue, had higher heritability, it is more likely that there is detectable genetic variation segregating for that cue’s role in flowering. Ideally, heritability when defining flowering using a specific cue should be higher than heritability using Julian date; otherwise, it makes little sense to use the cue for genomic prediction. We looked at heritability for nine traits associated with 50% flowering: daylength (analogous for a critical daylength for flowering), Julian date, cumulative GDD between greenup and flowering, the change in daylength from the previous day at flowering, and five measures of cumulative rainfall: cumulative rainfall between greenup and flowering, and in the two days, three days, five days, and seven days before flowering.

Because switchgrass phenology is closely correlated with growing degree days (GDD) during vegetative growth, we tested for a temperature-based environmental cue for flowering. To do this, we evaluated flowering date as a function of cumulative GDD between plant green up and flowering, using a base temperature of 12 C (Kiniry et al 2005, Behrman 2013). We then modelled GDD as a function of subpopulation and the interaction between subpopulation and latitude of origin, both as random effects. To avoid confounding with a photoperiod-based environmental cue, we removed GDD response values for genotypes predicted to have a photoperiod cue at the four common gardens south of 38°N. Including these GDD response values substantially increased the residual variance for photoperiod sensitive individuals at these sites (data not shown//in supplement). Subpopulation explained most of the variation in flowering as a function of GDD (% Varsubpop = 85.9%). Predicted values of GDD necessary for flowering were largest for the Gulf subpopulation (1056; 95% CI 973-1099), smallest for the Midwest (667; 95% CI 517 – 814) and intermediate for the other three groups (741 (599-1019); 776 (483-1195); 747 (573 – 1043); Supplementary table of effects). The heritability for GDD at flowering for photoperiod insensitive, sequenced individuals was 30%, while the heritability for flowering as a Julian date for the same set of individuals was 10.8% (Supplementary Table: Variance Components analysis). Thus, more phenotypic variance was explained by GDD than by Julian date for these individuals.

*Genome-wide association mapping*

We used the switchgrassGWAS R package (https://github.com/Alice-MacQueen/switchgrassGWAS) to allow fast, less memory intensive GWAS on the diversity panel.

*Analysis of correlated SNP effects on phenotypes at multiple sites*

We used the switchgrassGWAS R package to estimate and test the significance of SNP effects on phenology phenotypes measured at our common garden sites.

***Fourway cross and Quantitative Trait Locus Dataset***

*Creation, propagation, cultivation, and phenotyping*

A four-way population (801 F2 individuals) with their grandparents and F1 hybrid parents was clonally propagated 10 times in 3.8L pots at the Brackenridge Field Laboratory, Austin, TX in 2014-2015 and transplanted to 10 common garden sites covering 17 degrees of latitude from South Texas to South Dakota in May-July of 2015. The formation of the fourway mapping population has been described previously (Milano et al 2016). Briefly, the population was developed by initial cross between AP13 (A) x DAC6 (B) and WBC3 (C) x VS16 (D), the F1 hybrids of each of those crosses were then intercrossed reciprocally to create the four-way outbred population, 801 F2 individuals. The four grandparents were derived from highly divergent southern lowland and northern upland ecotypes: AP13, an early flowering Gulf individual, and WBC, a late flowering Gulf individual, DAC, an early flowering Midwest individual, VS16, a late flowering Midwest individual. The details of the 10 common garden sites can be found in Lowry et al. (2019). Briefly, each field was covered with one layer of weed barrier cloth, and holes were cut into the weed cloth in a honeycomb design. Plants were randomized into the holes, with each plant having four nearest neighbors located 1.56m away from each other. A row of border plants which were derived from rhizome plugs of approximately 10-year-old Alamo switchgrass were planted at every edge position to prevent edge effects. Plants were well-watered in 2015 to facilitate establishment.

Phenology data, including greenup and flowering time, have been recorded from 2016 to 2019 for the four-way population at each site. Greenup (GR50) was recorded as the day of the year when 50% of the tillers from the crown on the plant turned green, flowering (FL50) was recorded as the day of the year when 50% of the plant tillers had panicles undergoing anthesis. Additionally, cumulative growing degree days (GDD) from GR50 to FL50 was calculated as GDD = , where Tmean is the daily average temperature, Tbase is the base temperature of 12 °C for switchgrass (Kiniry et al 2005, Behrman 2013). On a specific day, if Tmean is less than Tbase, the GDD for that day is 0; if Tmean is bigger than Tbase, the GDD for that day is the difference between Tmean and Tbase. To be comparable and consistent with the diversity panel, only 2019 phenology data of the four-way population from the same seven common garden sites were used in this study.

*Sequencing and QTL mapping*

Illumina fragment paired end libraries, representing each of the four grandparents (A: AP13, B: DAC; C: WBC; D: VS16) were aligned to the *Panicum virgatum* reference genome v5) with bwa *mem* (Li and Durbin 2009). Details on the genetic map construction, map polishing and fine-scale reordering can be accessed on <https://datadryad.org/stash/dataset/doi:10.5061/dryad.ghx3ffbjv>.

Narrow-sense heritability (h2) for each phenotype (GR50, FL50, GDD) at each field site was estimated using the additive kinship matrix based on marker genotypic information. The process was accomplished via the ‘sommer’ package (Covarrubias-Pazaran, 2020) in R (2020). Briefly, we used a multivariate mixed model (mmer) that takes the kinship matrix and other random incidence matrices to estimate the variance components for each phenotype at each field, and calculates h2 as the proportion of additive genetic variance to the total variance.

QTL mapping was conducted with R/qtl2 (Broman, 2020). Specifically, we performed a genome scan with a linear mixed model accounting for the relationships among individuals (i.e., kinship matrix) and the environmental covariates (i.e., field sites). The full model can be expressed as: phenotype = µ + G + E + G x E + e, where µ is the population mean, G is the genetic effect (i.e., kinship matrix), E is the environmental effects (i.e., field sites), G x E is the interaction between genetic and environmental effects, and e is the error term. The genome scan was accomplished through ‘scan1’ function. The statistical significance of the genome scan was established by performing a stratified (i.e., stratifying on field sites) permutation test (n=1000) using ‘scan1perm’ function. The estimated QTL effect was obtained using ‘scan1coef’ function in R/qtl2.

***Data availability statement***

**References**

Parrish DJ, Fike JH. 2005. The Biology and Agronomy of Switchgrass for Biofuels. Critical Reviews in Plant Sciences 24:423-459.

Tables & Figures, Maybe Supplementary

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Site | Latitude | Genotypes that flowered while: | 4X | 8X | Atlantic | Gulf | Midwest |
| Kingsville, TX, USDA/PMC | 27.55 | days lengthening | 9 | 8 | 44 | 23 | 23 |
| Kingsville, TX, USDA/PMC | 27.55 | days shortening | 11 | 4 | 18 | 134 | 1 |
| Austin, TX, UT PRC | 30.38 | days lengthening | 19 | 167 | 206 | 11 | 115 |
| Austin, TX, UT PRC | 30.38 | days shortening | 12 | 66 | 56 | 173 | 6 |
| Temple, TX, USDA/ARS | 31.04 | days lengthening | 8 | 7 | 45 | 1 | 46 |
| Temple, TX, USDA/ARS | 31.04 | days shortening | 15 | 6 | 62 | 193 | 15 |
| Stillwater, OK, OSU | 35.99 | days lengthening | 1 | 0 | 2 | 0 | 2 |
| Stillwater, OK, OSU | 35.99 | days shortening | 18 | 8 | 86 | 83 | 52 |
| Columbia, MO, MU BRF | 38.90 | days lengthening | 0 | 3 | 0 | 0 | 6 |
| Columbia, MO, MU BRF | 38.90 | days shortening | 31 | 206 | 228 | 149 | 113 |
| Lincoln, NE, UNL ARF | 41.15 | days shortening | 18 | 13 | 123 | 66 | 67 |
| Hickory Corners, MI, KBS | 42.42 | days lengthening | 0 | 1 | 0 | 0 | 0 |
| Hickory Corners, MI, KBS | 42.42 | days shortening | 36 | 181 | 199 | 198 | 109 |
| Brookings, SD, SDSU | 44.31 | days shortening | 15 | 12 | 109 | 73 | 63 |

**Supplementary Figure X.** Average flowering date from 2016 – 2019 for the four parents of the fourway cross, compared to the 2019 distribution of flowering date for the Gulf and Midwest subpopulations. The two Midwest parents are DAC and VS16, and the two Gulf parents are AP13 and WBC. Bold yellow line indicates the summer solstice.

A close up of a map

Description automatically generated

A close up of a logo

Description automatically generated

However, cumulative GDD was not sufficient to explain the variation in flowering time in the Gulf subpopulation. Instead, we hypothesized that the Gulf subpopulation also used a photoperiod cue, given the imperfect relationship between GDD, photothermal time, and flowering in the Gulf subpopulation, and the presence of a relationship between latitude of common A screenshot of a cell phone

Description automatically generatedA close up of a map

Description automatically generatedgarden and flowering time based on cumulative GDD.

Cumulative GDD vs photoperiodicity as a function of each plant’s location of origin.