Statistical Analysis

Plants sown in 1996 were used in the analysis because it is the oldest cohort in the common garden experimental plot and would have the largest numbers of flowering plants during the study period (Table 1, Figure 1). This cohort contains plants from seven different remnants (Table 2). If a plant had multiple heads I took the earliest start date per head and latest end date per head as the plant's flowering period. I used an individual's first flowering day (FFD) as my primary way to assess differences and consistency in flowering time. FFD was highly correlated with an individuals end date within years (Pearson's correlation test; $r = 0.86$, $p < 0.001$).

Table 1: Number of flowering plants per year

year	n
2005	136
2006	202
2007	219
2008	160
2009	169
2010	117
2011	203
2012	51
2013	110
2014	93
2015	135

Table 2: Number of plants from each origin site

OriginSite	n
aa	33
eri	102
lf	66
ness	30
nwlf	161
spp	82
sap	96
Tower	4

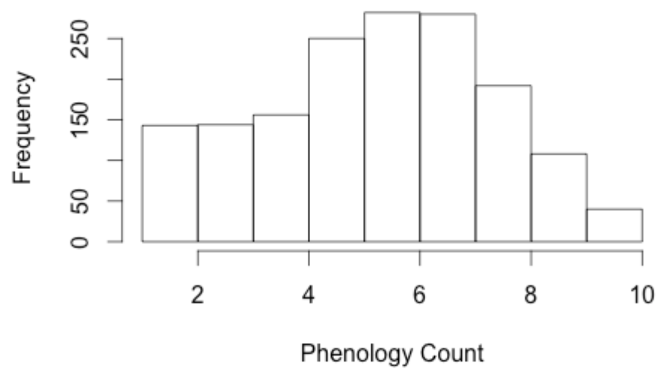

Figure 1: Number of times individual plants flowered over the study period (2015-2015)

Each plants first flowering day (FFD) was examined in three ways to assess the most appropriate manner to account for year-to-year differences in growing season time and duration. Day after first (DAF) is the number of days a plant began flowering after the first plant in that given year began flowering. A DAF of 1 signified an individual was the first plant to flower and a DAF of 10 signified a plant began flowering 10 days later (Figure 2A). Day of year (DOY) is the date on which a plant began flowering and represents the data in its orginal form (Figure 2B). Day after median (DAM) ranked plants by median FFD in a given year. Individuals that began flowering on the same day as the median FFD received a rank of zero. Those that began flowering two days prior to the median FFD received a rank of -2 and those that began flowering 2 days after the median FFD received a rank of 2 (Figure 2C).

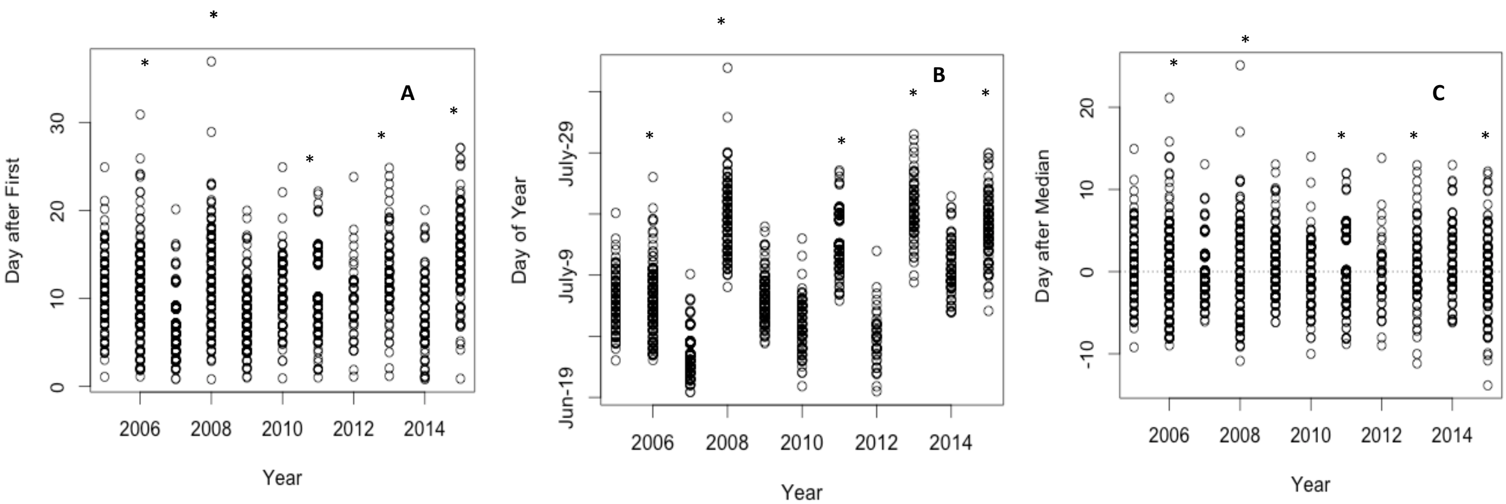

Figure 2. Distribution of FFDs of all plants in all study years. A) Flowering rank based on the first flowering individual in every year. B) FFDS as calendar day of year. C) FFD transformed to day after median score. Years when the experimental plot was burned are denoted with asterisks.

Flowering rank of individual plants were tested for association among years using pairwise Spearman rank correlations. Sample sizes for pairwise year comparisons varied considerably, yet nearly all years (except 2005 and 2012) had significant correlation between ranks of those plants that flowered in those two years (Table 1).

Table 3: P-values for pairwise Spearman rank correlations of first flowering days of individual plants in the given years

	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
2015											
2014	0.097										
2013	**0.015***	0.244									
2012	0.171	0.103	0.521								
2011	**<0.001***	**0.007***	**0.004***	0.258							
2010	0.063	**0.045***	0.172	0.362	0.732						
2009	0.270	**0.016***	0.141	0.648	0.086	**0.006***					
2008	**0.025***	0.886	**0.003***	0.185	**0.009***	**0.004***	0.948				
2007	**0.015***	**0.016***	**0.010***	0.568	**<0.001***	**0.034***	**<0.001***	**<0.001***			
2006	**0.024***	0.999	0.092	0.101	**0.009***	**0.015***	**0.0467***	**<0.001***	**<0.001***		
2005	0.181	0.274	0.153	0.167	0.994	0.215	0.305	0.478	0.248	0.064	

To assess whether individuals consistently flower at the same time relative to each other I created a null model that, for every plant that flowered in a given year, reassigned DAM values based on all the dates in each year. These DAM values were reassigned at random, with replacement, for all years. This model shuffled dates that individuals flowered in a given year, but retained individual interannual variation in flowering time. I resampled the model 10,000 times and excluded plants that only flowered once during the study period. I tested two parameters from this model – mean DAM and range of DAM. Mean DAM is an individuals average DAM score across the study peroid and range of DAM was calculated by the range in each individuals FFD across the study period. If individuals are consistent in their FFD across years, then the variance of the mean DAM values of the observed data would be greater than the variances obtained through resampling the null model. Likewise, if individuals are consistent than the variance of the range of DAM values (the difference between an individuals min and max DAM) of the observed data would be smaller than the variances obtained through resampling the null model.

The variance of mean DAM ranks was significant in comparison to the null model (bootstrap, $p < 0.001$, Figure 3.). Higher variance means the observed data had a wider distribution, whereas the average variance in the bootstrap model was 4.35, significantly lower than the observed data, and implying that plants are more consistent in their flowering than just by random chance.

I broke this model down further and examined the middle third of plants, those that have a mean DAM rank between -1 and 1. Plants with a central mean could either been consistently flowering in the middle of the flowering season, or the average could masking both early and late start dates (inconsistent flowering) (Figure 4).

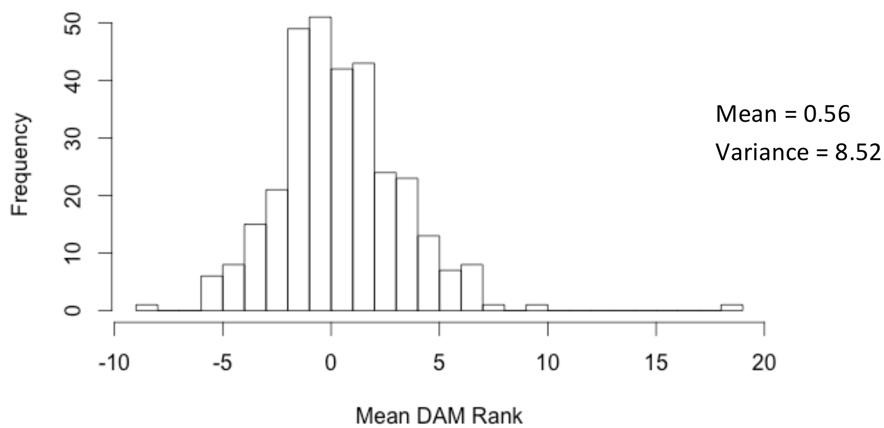

Figure 3. Histogram of the mean Day after Median (DAM) rank for all plants in the study period. Plants that only flowered once were excluded from the analysis.

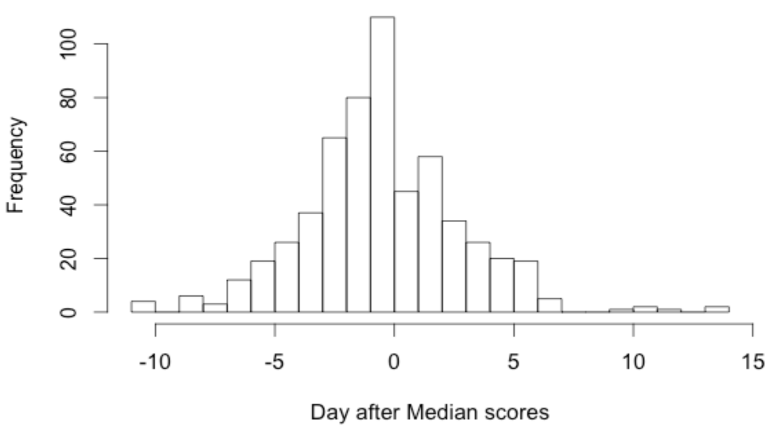

Figure 4. Day after median flowering scores for plants with an average DAM between -1 and 1

When the same bootstrap analysis is run on those plants with a mean DAM between -1 and 1, the average variance of the model (var = 2.5) is significantly higher than our observed data (var = 0.49, $p<0.001$). This implies that the plants with middle averages are not consistently flowering in the middle of the growing season.

Lastly, if the same boostrap analysis is run on the data excluding the middle third, then we end up with similar results as if all the data are together. The observed variance (12.5) is significantly higher than the bootstrap resampling ($p <0.001$, $n = 10000$).

Plants in the experimental plot originated from 7 different remnants, however since there are only 4 plants from the remnant “Tower”, they were excluded from the analysis. Mean DAM rank varied between site and the number of times plants flowered across the study period (Figure 5).

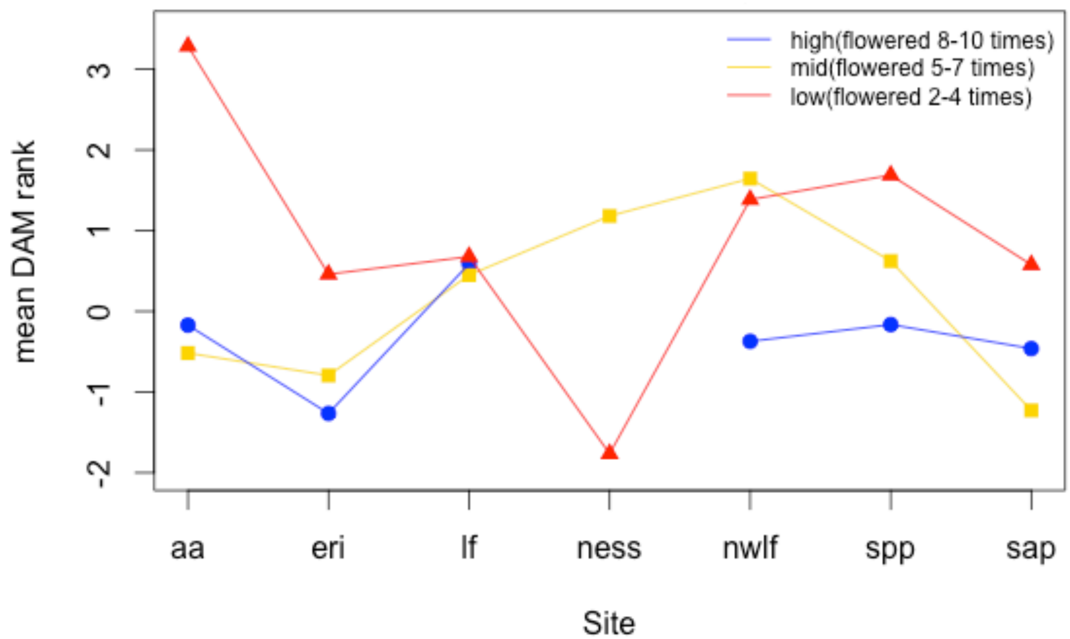

Figure 5. Average DAM for plants from each of the origin sites. Plants are further subdivided by number of times they flowered over the study period.

Simple backwards model elimination was used to test whether origin site mattered when predicting average DAM rank and annual DAM rank (Table 4). Site, the number of times a plant flowered over the study period, as well as its position in the experimental plot all mattered when testing models (Table 4).

Table 4: ANOVA table comparing models of average Day after Median rank. Models include the following variables: origin site (S), number of times the plant flowered during the study period (C), and their position in the experimental plot (R and P). Although the insignificant p-value for model 5 indications that row does not matter in predicting a plant’s mean DAM rank, it’s included in the model because a secondary analysis showed that when only row, pos, and their interaction are used to predict mean DAM rank, row is significant.

Model	Compared with	Focal term	Model DF	Model SS	Test DF	Test SS	Test F	Test *P*-value
1. S + C + S*C + R + P + R*P			1412	8274.2				
2. S + C + S*C + R + P	**1.**	**R x P**	**1413**	**8281.9**	**-1**	**-7.743**	**1.321**	**0.251**
3. S + C + R + P	2.	S x C	1419	8539.7	-6	-257.73	7.329	<0.001
4. S + C + S*C + R	2.	P	1413	8281.9	1	113.43	19.352	<0.001
5. S + C + S*C + P	2.	R	1413	8281.9	1	12.61	2.152	0.143

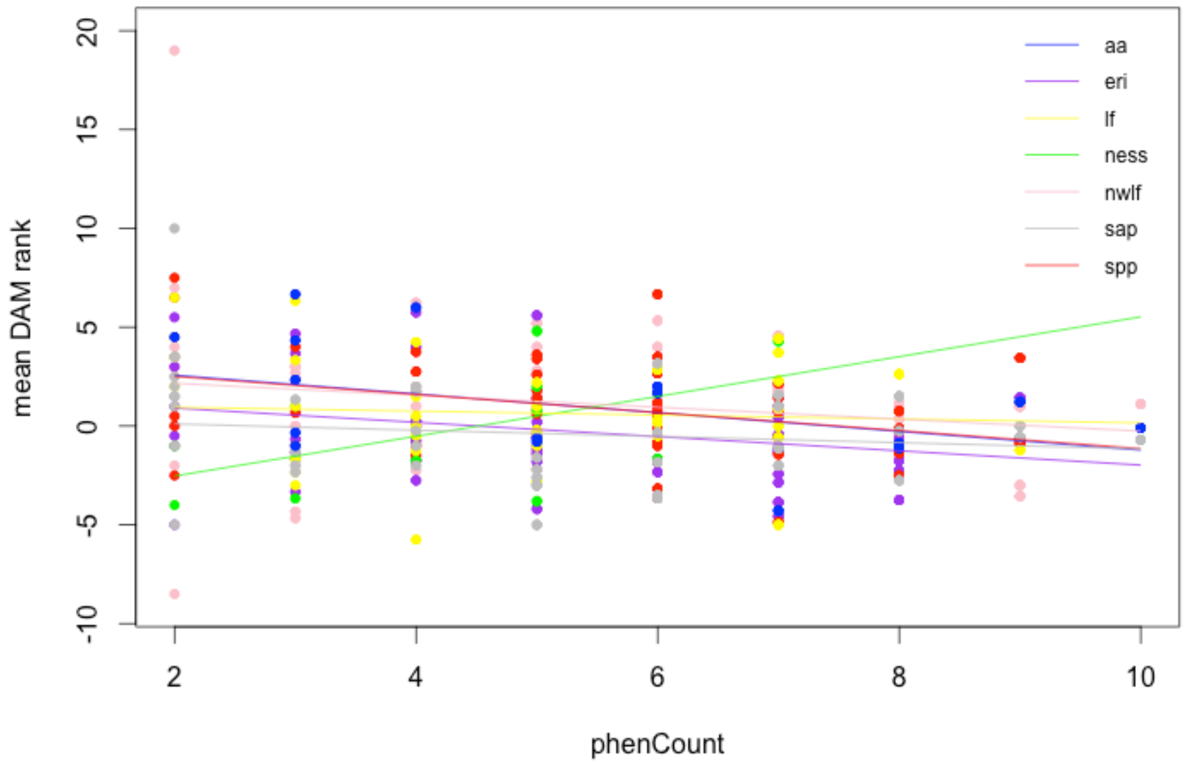

Figure 6. Mean Day after Median scores of plants by the number of times they flowered over the study period (phenCount). Colors indicate origin site and lines are predicted values of mean Day after Median scores for each site at each phenology count.

The observed variance of the range of DAM was only marginally significant in comparison to the bootstrap resampling model (bootstrap, n = 10000, p = 0.06, Figure 7). The average variance in the model was 21.96.

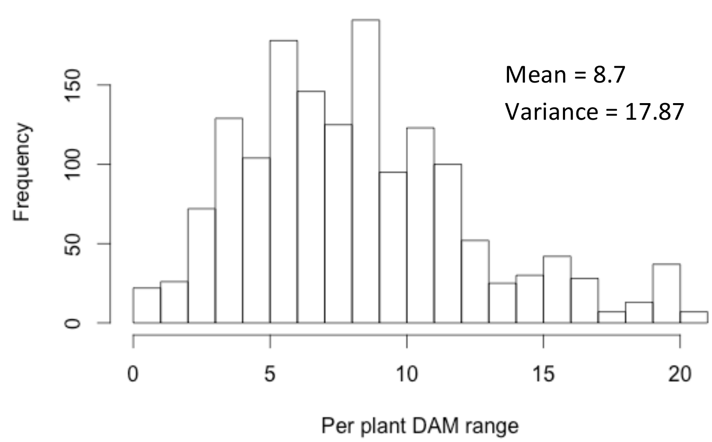

Figure 7. Histogram of the Day after Median ranges for all plants in the study period.

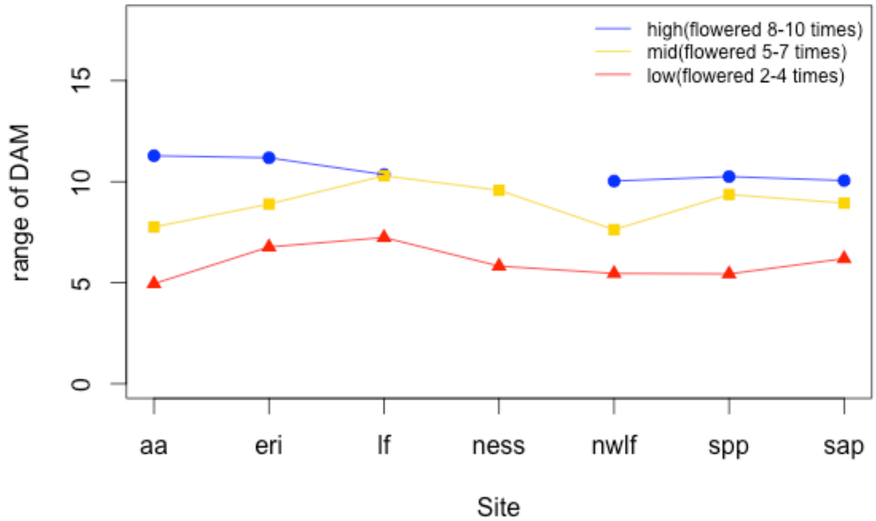

Figure 8. Range of DAM for plants from each of the origin sites. Plants are further subdivided by the number of times they flowered over the study period.

Simple backwards model elimination showed that site (S), the number of times a plant flowered over the study period (C), as well as position in the experimental plot (R and P) mattered when predicting the range of a plants' Day after Median score (Table 5).

Table 5: ANOVA table comparing models of range in Day after Median rank. Models include the following variables: origin site (S), number of times the plant flowered during the study period (C), and their position in the experimental plot (R and P).

Model	Compared with	Focal term	Model DF	Model SS	Test DF	Test SS	Test F	Test *P*-value
1. S + C + S*C + R + P + R*P			1412	18134				
2. S + C + S*C + R + P	**1.**	**R x P**	**1413**	**18143**	**-1**	**-8.3274**	**0.6484**	**0.4208**
3. S + C + R + P	2.	S x C	1419	18377	-6	-234.28	3.041	0.0059
4. S + C + S*C + R	2.	P	1414	18469	-1	-326.1	25.398	<0.001
5. S + C + S*C + P	2.	R	1414	18210	-1	-67.126	5.2279	0.0224

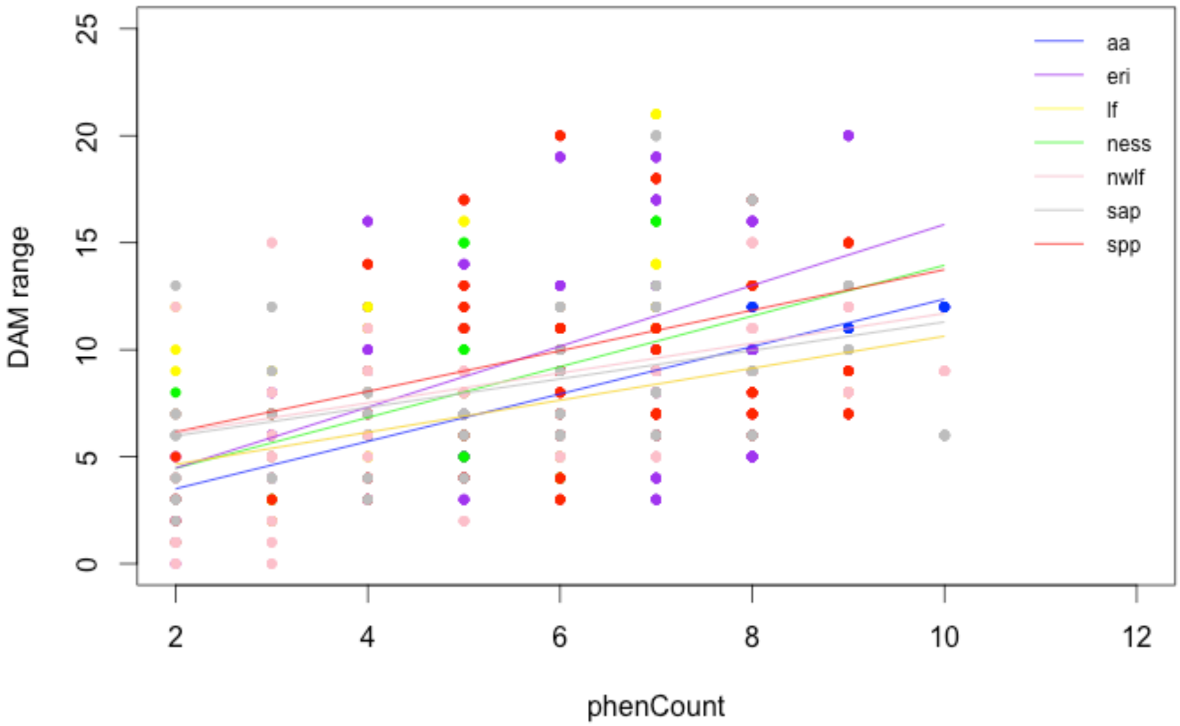

Figure 9. Range of Day after Median scores of plants by the number of times they flowered over the study period (phenCount). Colors indicate origin site and lines are predicted values of the range of Day after Median scores for each site at each phenology count.