2021

Genome Wide Association Study on Kernel Size in *Avena Sativa* (Oats)

Koryne Ann Carlson
South Dakota State University

GENOME WIDE ASSOCIATION STUDY ON KERNEL SIZE IN *AVENA SATIVA* (OATS)

BY

KORYNE CARLSON

A thesis submitted in partial fulfillment of the requirement for the

Master of Science

Major in Plant Science

South Dakota State University

2021

THESIS ACCEPTANCE PAGE

Koryne Carlson

This thesis is approved as a creditable and independent investigation by a candidate for the master's degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Melanie Caffe
Advisor Date

David Wright
Department Head Date

Nicole Lounsbery, PhD
Director, Graduate School Date

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Melanie Caffe for being willing to be my thesis advisor and taking me on as a student. I have learned a lot from her and her program. I appreciate her willingness to work with me as I am not a traditional master's student and worked while obtaining my master's degree. Her guidance on how research is conducted has been very beneficial for me to learn.

I would like to thank General Mills for sponsoring me and my project. Without them I would not be able to get my master's degree. I would also like to thank those at General Mills for their guidance and willingness to let me obtain my M.S. degree. I would especially like to thank Paul Richter for pushing me to get my master's degree, Tom Rabaey and Julie Holmstrom for being amazing managers and approving all of my class requests, and Steve Eichten for helping guide me during the association mapping portion of my project.

I would also like to thank my family for supporting and encouraging me to pursue my master's degree. Without their push I would not be where I am today.

TABLE OF CONTENT

ABBREVIATIONS

SDSU	South Dakota State University
CPT	Crop Performance Testing
VIBE	Qualmaster Computer Vision Device
LSD	Least Significant Difference
DLRF	Dakota Lakes Research Farm
SERF	Southeast Research Farm
NERF	Northeast Research Farm
QTL	Quantitative Trait Loci
PYT	Preliminary Yield Trial
GWAS	Genome wide association study
GBS	Genotype by sequencing
LD	Linkage disequilibrium
PCA	Principal component analysis
LOD	Logarithm of odds
SNP	Single nucleotide polymorphism

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

GENOME WIDE ASSOCIATION STUDY ON KERNEL SIZE IN *AVENA SATIVA* (OATS)

KORYNE CARLSON

2021

Kernel size is a vital milling characteristic for *Avena sativa* L. (oats), with larger kernel size being desirable for milling facilities. This project evaluated the effect of genotype, environment, and their genotype by environment interaction (G by E) on kernel width, length, and area for a set of 27 oat cultivars grown at multiple locations in South Dakota over three years. Findings indicated significant differences among oat varieties, and that environment had a substantial effect on kernel size. However, we did not see a strong influence of genotype by environment interaction on kernel size. The lack of impact for genotype by environment interaction indicates that the rank order of the varieties will be similar from one environment to another, meaning kernel size measurement can occur in fewer locations and still successfully identify genotypes with larger width, length, and area. Broad-sense heritability was also evaluated. Our long-term objective is to develop new oat varieties with larger kernel size. With broad-sense heritability estimates ranging between 0.42 – 0.79 in our study, breeding for increased kernel size will be achievable. We did a genome-wide association study (GWAS) to locate the significant quantitative trait loci (QTL) that control kernel size. Using the recently published oat genome, we found four loci for kernel width and length as well as the RNA sequences from the reference genome. Only one of those RNA sequences had high similarity with a sequence

(TRINITY_DN32700_c0_g1_i1) found in Emmer. By finding the markers that impact kernel size marker assisted selection can now be used to help breeders breed for larger kernel size.

CHAPTER 1: EFFECT OF GENOTYPE, ENVIRONMENT, AND GENOTYPE BY ENVIRONMENT INTERACTION ON KERNEL SIZE IN OATS

ABSTRACT

Kernel size in *Avena sativa* L. (oats) is vital for milling facilities for their cleaning process of contaminant seed for cleaner purer oats. We studied three years of kernel size (width, length, and area) data from the South Dakota Crop Performance Trial. Twenty-seven or twenty-four varieties were evaluated at six to eight locations in South Dakota from 2018 to 2020. Thirteen varieties were evaluated in all three years of evaluation, representing a total of 21 environments. Differences in kernel size among oat varieties were identified. There is not a significant effect of genotype by environment interaction on kernel size. The rank change among genotypes for kernel size from one environment to another should therefore be minimal. It also suggests that we can measure kernel size in fewer environments. Broad-sense heritability was calculated using the variance components. It was found that kernel width, length, and area have relatively high heritability, with broad-sense heritability estimates ranging between 0.43-0.69, 0.5-0.79, 0.66-0.79 for width, length, and area, respectively. It was also discovered that there is not a strong correlation between nutritional characteristics and kernel size.

1.1 INTRODUCTION

Oats (*Avena sativa L.*) are ranked sixth in world grain production (Francis H Webster, 2011). The majority of the oat acres grown are harvested for forage or animal feed and cover crop, while a small amount goes to human consumption in the U.S. Over the last 50 years, U.S. oat acres have drastically declined by 60% because of farming's mechanization (Francis H Webster, 2011). About 10% of the world's oat crop is used for food (Francis H Webster, 2011). Oats are known for having beta-glucan, the heart-healthy fiber that helps lower cholesterol. For this reason, oats are widely consumed by humans for health benefits. Oats are the highest-protein cereal grain crop (Francis H Webster, 2011).

The oat plant has a panicle which is where the seed is produced at the top portion of the plant. The panicle is made up of many florets (a small flower that grows the seed). The oat spikelet is made up of the primary, secondary, and tertiary florets which can each bear a seed. According to Doehlert, the largest kernel type is the primary seed (Douglas C. Doehlert M. S., 2002) and, tertiary is the smallest kernels found on the oat plant. Tertiary kernels are the last of the seeds to develop, and some oat varieties do not produce a tertiary kernel. Doehlert also found that kernels closer to the top of the panicle are larger (Douglas C. Doehlert M. S., 2002). Kernel size distribution is bimodal due to different kinds of kernels found on the oat plant, primary, secondary and tertiary (Douglas C. Doehlert M. S.-L., 2004). Varieties with higher uniformity for kernel size also have larger kernel size (Douglas C. Doehlert M. S.-L., 2004).

A larger kernel size is desirable because it helps the milling facilities with the separation of contaminant seed. Unlike wheat, barley, and rye, oat does not contain

gluten. For milling facilities to keep gluten out of their products the contaminant seeds need to be sorted out. The separation process to remove contaminant seed from desired seed is based on size. Oat kernels tend to be larger than wheat, barley, and rye, in part because the hull remains attached to the oat groat following threshing. Small-sized oat kernels cannot be separated from the wheat, barley, and rye and are therefore discarded. Larger kernels are also beneficial because they make larger flakes. Oat flakes are formed when the groats (oats without the hull) are rolled flat. With the larger flakes the oats can be made into more products. It has been found that smaller oat kernels require more energy to dehull (Doehlert 2004). Because of this, larger kernels will require less energy to dehull saving the milling company time and energy.

For this project, we wanted to determine if there were statistical differences between kernel sizes in varieties, and how much the environment affects width, length, and area. Our objective was also to estimate the heritability of kernel width, length and area to make sure that breeders can make gains when breeding for larger kernel sizes. It is also essential to determine if any correlation between kernel size and nutritional quality characteristics exist and ensure that these traits are heritable. It was previously found that beta glucan is found in the aleurone and sub aleurone (outer layers) layers of the kernel (Per Sikora, 2013), indicating that kernel size impacts the amount of beta glucan in the kernel. So, the larger the kernel the lower the beta glucan.

1.2 MATERIALS AND METHODS

1.2.1 Sample composition

Oat grain samples from the SDSU Crop Performance Testing (CPT) oat variety trials conducted in 2018, 2019, and 2020 were used for this study. Twenty-seven lines in

2018, 2019, and twenty-four lines in 2020 were tested. The entries are listed in Table 1. Entries consisted of commercially available varieties and SDSU experimental lines. The CPT trials are grown at multiple locations in common oat-growing regions of South Dakota. Grain samples used for this study were harvested from 6, 7, and 8 locations across the state in 2018, 2019, and 2020, respectively (Table 2). A map of the locations can be found in Figure 1. At each location, the field experiment consisted of a randomized complete block design with four replications. The first two replications were used for this study. A total of 13 oat entries were common in all three years and evaluated in twenty-one environments (year/location combinations).

1.2.2 Kernel size measurement

Kernel size was measured using Qualmaster Computer Vision Device (VIBE) (VIBE Technologies, Tel Aviv, Israel). This high throughput device (Figure 2) measures kernel size: length, width, and area, as well as color; we did not use the color output for our analysis. The VIBE works by capturing an image of a seed sample with the camera's geometric calibrated lens. Once the picture is captured, the background is removed using a segmented color technique, and each kernel is then characterized as a shape on a binary image. A minimal bounding rectangle is used to extract the shape for measuring length, width, and area for each kernel (Moran Nave R. A.-Z., 2016). For our analysis, a 10g sub-sample was used to measure the width, length, and area. The kernels were scattered so that they were not touching to ensure measurement accuracy. If the kernels overlap, the instrument cannot identify each kernel individually and instead identifies the overlapping kernels as one. The output is provided as a .csv file and includes measurements for each kernel and the sample average. A calibration was developed so

that any kernel 25% smaller than the sample average was considered broken. Based on this broken threshold, the output included a sample average and whole kernel average. The whole kernel average only factored whole kernel measurements and was used for the analysis, so broken kernels did not skew our measurements.

1.2.3 Data analysis

To test for a normal distribution a Shapiro-Wilk normality test was done on each year for length, width, and area. Our hypothesis for the Shapiro-Wilk test was Ho: the population is normally distributed and H1: the population is not normally distributed.

A Pearson correlation was performed looking at the correlation between width, length, and area to nutritional and quality characteristics. This correlation is vital because we do not want larger kernel size to impact oat quality and nutritional characteristics.

Data analysis was performed using the minque and agricolae packages in R (R Core Team 2020). Frequency histograms were created to identify and remove outliers. Samples with measurements outside the threshold of 2.25 – 3.5 mm for width and 8 – 13mm for length were considered an outlier. Since the area is dependent on length and width, we did not have an outlier threshold for the area. An ANOVA was performed to determine a statistical difference among the different varieties and the different locations. The linear mixed model that was used for the ANOVA, the model was $y_i = \mu + v_j + e_k + ve_{jk} + \varepsilon_l$. Where μ is the overall mean, v is variety, e is the environment, ve is the interaction between variety and the environment and ε is random error.

Based on the ANOVA results, a least significant difference (LSD) model was performed. The equation for LSD is $LSD = t\sqrt{MSW(\frac{1}{n_a} + \frac{1}{n_b})}$. Where t is the critical

value from the t distribution table, MSW is the mean square within from the ANOVA table and n is the number of varieties used. We used the package agricolae (Mendiburu, 2020) in R to perform the LSD. The variance components were calculated with the R package minque (Wu, 2019) using the same linear mixed model as above; all terms are considered to be random. The linear mixed model used the jack-knife resampling approach. After the variance components were calculated, broad-sense heritability could be estimated using the following calculation: $H^2 = \frac{V_g}{V_p}$ where V_g is the genetic variation and V_p is the phenotypic variation which is composed of the genetic and environmental variation.

1.3 Results

1.3.1 Kernel width

For the shapiro-wilk normality test 2019, 2020 and 2018-2020 were all normally distributed, while 2018 was not a normal distribution.

Since our breeding objectives are to develop varieties with larger kernel size, the oats' nutritional quality mustn't be affected by the difference in width, length, and area. Width is positively correlated with the percentage of plump kernels (r = 0.68). The percentage of plump kernels is measured by shaking a 100g sample over a 5.5/64" sieve for 30 strokes. This correlation makes sense when examining kernel width. Kernels with a width higher than 5.5/64" are retained on top of the sieve. Width was strongly negatively correlated with the mid-measurement (r = -0.64). The mid-measurement corresponds to kernels that fall through the 5.5/64" sieve but are retained on top of the 5/64" sieve. Nutritional quality characteristics (beta-glucan, protein, and oil content) were

not significantly correlated with width, with beta glucan, protein and oil having an r value of -0.14, -0.09, and -0.13, respectively (Table 11). This is important because there has been significant breeding effort to have higher beta glucan and protein and lower oil content. The new effort to breed for larger kernel size the nutritional content will not be affected.

Histograms for each year of the width distributions are in Figure 3, and the histogram for the combined years is in Figure 6. The average whole kernel width of 543 oat grain samples grown at multiple locations in South Dakota from 2018 to 2020 ranged from 2.65 to 3.12 mm with an average of 2.84 mm. An analysis of variance was done for data collected each year and for the combined years. Those analyses revealed that the effect of genotype, environment, and year was significant. However, the interactions between genotype and environment on kernel width were not significant. Genotypes represented the largest source of variation (57.7 and 46%) for kernel width in 2018 and 2019 and the second-largest source of variation (35.5%) in 2020 (Table 9). The average kernel width for every variety tested is in Table 3. Experimental line SD140327 was only evaluated in 2018 but produced kernels with the largest average width (3.12 mm) that year. In 2019, SD120665 produced kernels with the highest average width (3.06 mm). In 2020, SD160070 produced kernels with the largest average width of 3.00 mm; SD160070 was only evaluated in 2020. On the other hand, genotypes Antigo (2.67 and 2.66 mm in 2018 and 2020, respectively) and SD160201 (2.65 mm) in 2019 produced seed with the smallest average width. Broad-sense heritability estimate for width ranged from 0.43 to 0.69 (Table 10), suggesting that width has relatively high heritability and genetic gain through breeding efforts should be easily achievable.

Location was the second-largest source of variation (25.6% and 21.2%) in 2018 and 2019 and the largest source of variation (46.2%) in 2020. Grain with the largest average width (2.98 mm) was produced at the NERF in 2020 (Table 6). On the other hand, grain with the smallest width (2.73 mm) was produced at Volga in 2018 and Dakota Lakes Research Farm (DLRF) in 2020. Volga is a location that is known for a lot of disease pressure, while DLRF is a dry environment. That could be a contributing factor for the smaller kernel size.

1.3.2 Kernel length

The Shapiro-Wilk normality test showed that all length populations were normally distributed.

Whole kernel length was positively correlated with thousand kernel weight (r= 0.38) (Table 11). The longer the kernels, the heavier they are, contributing to a larger thousand kernel weight. There was not a strong correlation for any of the nutritional characteristics, with beta glucan being r=0.23, protein r=-0.27 and oil r=-0.21 (Table 11).

Histograms for kernel length for each variety for all three years of the study can be found in Figure 4, and the length histogram for all three years combined is in Figure 6. The average kernel length was measured for 543 oat grain samples grown at multiple South Dakota locations for 2018, 2019, and 2020. The average kernel length ranged from 9.38 – 12.22 mm with an overall average of 10.72 mm (Table 4). The ANOVA showed that the effect of genotype, location, and year was significant. The interactions between genotype and environment on kernel length were not significant. Genotype accounted for the most prominent source of variation (66.5, 70.3%) for 2018 and 2020 and similar (42.9%) to the environment in 2019 (Table 9). Variety CS Camden consistently produced

grains with the largest average length for each year, with the largest length being 12.22 mm in 2020 (Table 4). The shortest average length was consistently Antigo, with the smallest average length being 9.38 mm in 2020. The broad-sense heritability for length ranged from 0.5 – 0.79 (Table 10). Having high of a heritability estimate shows us that kernel length is a trait that breeders can make genetic gains on and length can be improved.

Location was the second-largest source of variation for kernel length at 17.9%, 42.8%, and 17.7% for 2018, 2019, and 2020, respectively (Table 9). The location with the largest average length was Aberdeen (11.30 mm) in 2020 (Table 7). Grain with the shortest average length was produced in Winner (10.10 mm) in 2019. Winner is one of the earlier locations planted but is often drought stressed due to its more wester climate, which may explain why grain produced at that location had in average shorter length.

1.3.3 Kernel area

The shapiro-wilk results for area are 2018 are normally distributed and 2019, 2020 and 2018-2020 are not normally distributed.

The correlation between area and the various milling characteristics and nutritional quality traits was evaluated. Kernel area was significantly correlated with thousand kernel weight (r=0.63) (Table 11). This makes sense because the larger the area, the larger the kernels' weight, leading to heavier kernels. The area was negatively correlated with thin (r=-0.44). The thin measurement is what passes through a 5/64” sieve in the shaker measurement described above. With increased efforts in breeding for larger kernel size, a negative correlation between nutritional quality and larger kernels could lead to reduced nutritional quality, it was found in 2013 that majority of the beta glucan is

found in the aleurone and subaleurone layers (Per Sikora, 2013). Since the aleurone layer is around the outside it has been thought that the larger the kernel size lowers the beta glucan content. From our correlations found in our population that is not the case. For area we had a correlation of r=0.10 for beta glucan, r=-0.18 for protein and -0.22 for oil (Table 11).

The average whole kernel area of 543 oat samples was measured. Histograms of the kernel area for each variety for all three years are shown in Figure 5, and the area histogram for all three years combined is found in Figure 6. The average area ranged from 17.62 to 23.96 mm^2 for the years 2018 to 2020 (Table 5). The ANOVA revealed that location, genotype, and year had a significant effect on kernel area. Yet, the interaction between environment and genotype was not significant, as observed for width and length. Genotypes represented the largest variation source in all three years of evaluation (52.8%, 50.1%, and 62.2%) (Table 9). The genotype with the largest average area was CS Camden (32.96 mm) in 2019, while the genotype with the smallest average kernel area was Antigo (17.62 mm) in 2020 (Table 5). The broad-sense heritability estimates for the area ranged between 0.66 – 0.79 (Table 10). The kernel area has a heritability estimates similar to the other two traits, with area being calculated using width and length, this is what we would expect to see. We can make genetic gains when breeding for larger area because it has high heritability.

Location was the second-largest source of variation (26.9%, 22.4% and 18.0%) (Table 9). The average kernel area for each environment ranged from 19.70 to 22.91 mm^2. Kernels with the largest area were produced at the NERF (22.91 mm^2), and kernels with the smallest area were produced at the Volga site (19.70 mm^2) (Table 8). The NERF

has good growing conditions with little disease pressure and is usually slightly cooler in temperature. Volga is heavy in disease pressure every year; this can account for the difference in kernel area between the two locations.

1.4 Discussion

Some of our populations were not normally distributed. It was found in 2002 that oat kernel sizes are multimodal because of the different kinds of kernels the oat plant produces (Doehlert, 2002). It was found that some of the populations used were multimodal and not normally distributed. The samples used in this analysis were randomly pulled from the bag, so a variety of kernel sizes were measured. Leading to a range of kernel sizes measured.

It was found that there is not strong correlation between kernel size, quality, and nutritional quality characteristics. This is a good thing, there has been great effort in breeding for improved nutritional quality (higher beta glucan and protein content; and lower oil content). We do not want breeding effort for larger kernelled oats to negatively impact milling and nutritional quality traits. There was a paper that found majority of the beta glucan in the outer layer of the oat kernel indicating that beta glucan is more prominent in smaller kernels (less surface area) (Per Sikora 2013). This led us to believe that the amount of beta glucan is correlated with kernel size. Here we found that is not the case. Beta glucan is not correlated with kernel size.

There was little genotype by environment interaction meaning fewer environments are needed to evaluate kernel size. The ranking among genotypes from the largest kernel size to the smallest kernel size will not change much between environments.

There will still be variation for kernel size among environments. When looking at the average kernel size, the largest kernel size environments are environments with ideal growing conditions. In contrast, locations with the smallest average kernel size had more disease pressure and were often dryer climates. It was found in 2002, that the genotype plays a major role in kernel size (Doehlert, 2002), which is what we found, for width, length, and area. Genotype played the largest source of variation for all of the length and area measurements and for 2018, and 2019 width. Doehlert et al 2002 also found the order that kernel size was affected: kernel type, genotype, environment and position on the panicle. This is what we found when looking at genotype and environment. We did not differentiate kernel type or position on the panicle in our analysis.

Despite our relatively small genotype set, kernel size related traits (width, length and area) had high heritability. In this population and set of environments, length and area had high heritability and width had moderate heritability. This suggests that breeders can make genetic gains when breeding for larger kernel size. They can use the largest kernelled varieties as parents to develop new varieties with large kernel.

There are some implications when it comes to this study. We only looked at a small population for kernel size. In future research, there should be a larger population used over multiple years. The genotypes we tested are all highly adapted to grow in South Dakota, they are not a diverse population. There should be a diversity panel used or more genotypes used from different breeding programs. The field trial locations could also be more diverse. We only looked at locations in South Dakota. Oats are grown in many more environments across the US.

TABLES AND FIGURES

Table 1. List of oat varieties evaluated in the Crop Performance Trials from 2018 to 2020.

2018	**2019**	**2020**
Antigo*	Antigo*	Antigo*
CS Camden*	CS Camden*	CS Camden*
Deon*	Deon*	Deon*
Goliath*	Goliath*	Esker2020
Hayden*	Hayden*	GM 2015Y3232
Horsepower	Horsepower	Goliath*
Jerry	Jerry	Hayden*
Jurry	Jurry	MN Perl
Natty*	MN Perl	Natty*
Newberg	Natty*	Rushmore*
Rockford	Newberg	Saddle*
Rushmore*	Rushmore*	Shelby427*
Saddle*	Saddle*	Warrior*
Shelby427*	Shelby427*	SD140741*
Souris	Sumo	SD150012*
Sumo	Warrior*	SD150270*
Warrior*	SD140741*	SD160070
SD120665	SD120665	SD160071
SD140313	Rushmore*	SD160201
SD140327	SD150015	SD160240
SD140517	SD150020	SD170463
SD140741*	SD150081	SD170467
SD141043	SD150112	SD170479
SD141130	SD150270*	SD170970
SD150012*	SD160240	
SD150015	SD160567	
SD150270*	SD160778	
*designates in all 3 years		

Table 2. List of SD Crop Performance Testing Oat Variety Trial locations for 2018, 2019, and 2020.

2018	2019	2020
Aberdeen	Aberdeen	Aberdeen
Miller	Dakota Lakes	Dakota Lakes Research Farm (DLRF)
Northeast Research Farm (NERF)	Miller	Miller
Selby	Northeast Research Farm NERF	Northeast Research Farm (NERF)
Volga	Southeast Research Farm (SERF)	Selby
Winner	Volga	Southeast Research Farm (SERF)
	Winner	Volga
		Winner

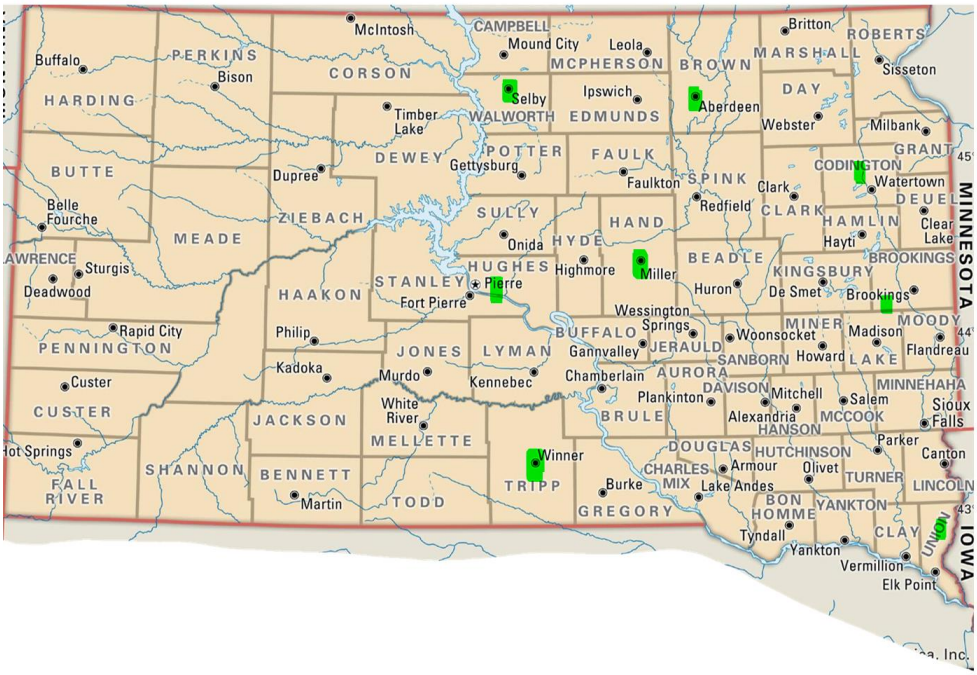


Figure 1. The locations used for this study is highlighted in green on the map of South Dakota.

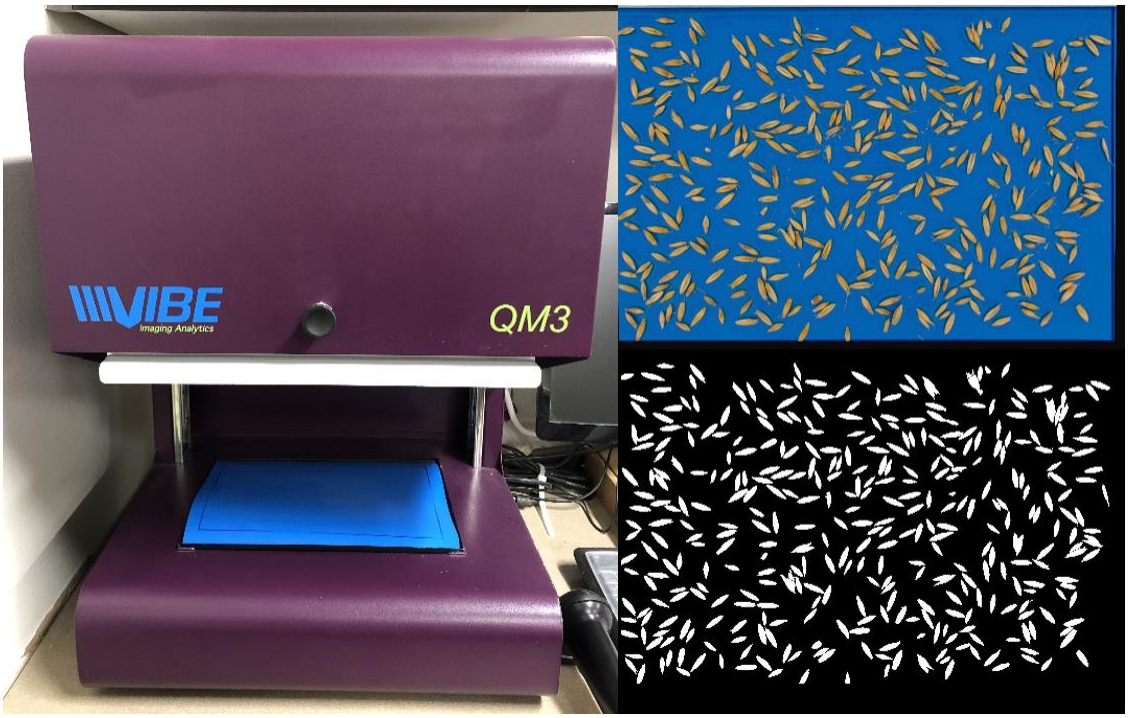


Figure 2. VIBE with a sample ready to be analyzed. The left picture is of the instrument itself. The top right is of a sample spread out on the mat, ready to be analyzed. The bottom right is the binary image used for the kernel size measurements of the sample.

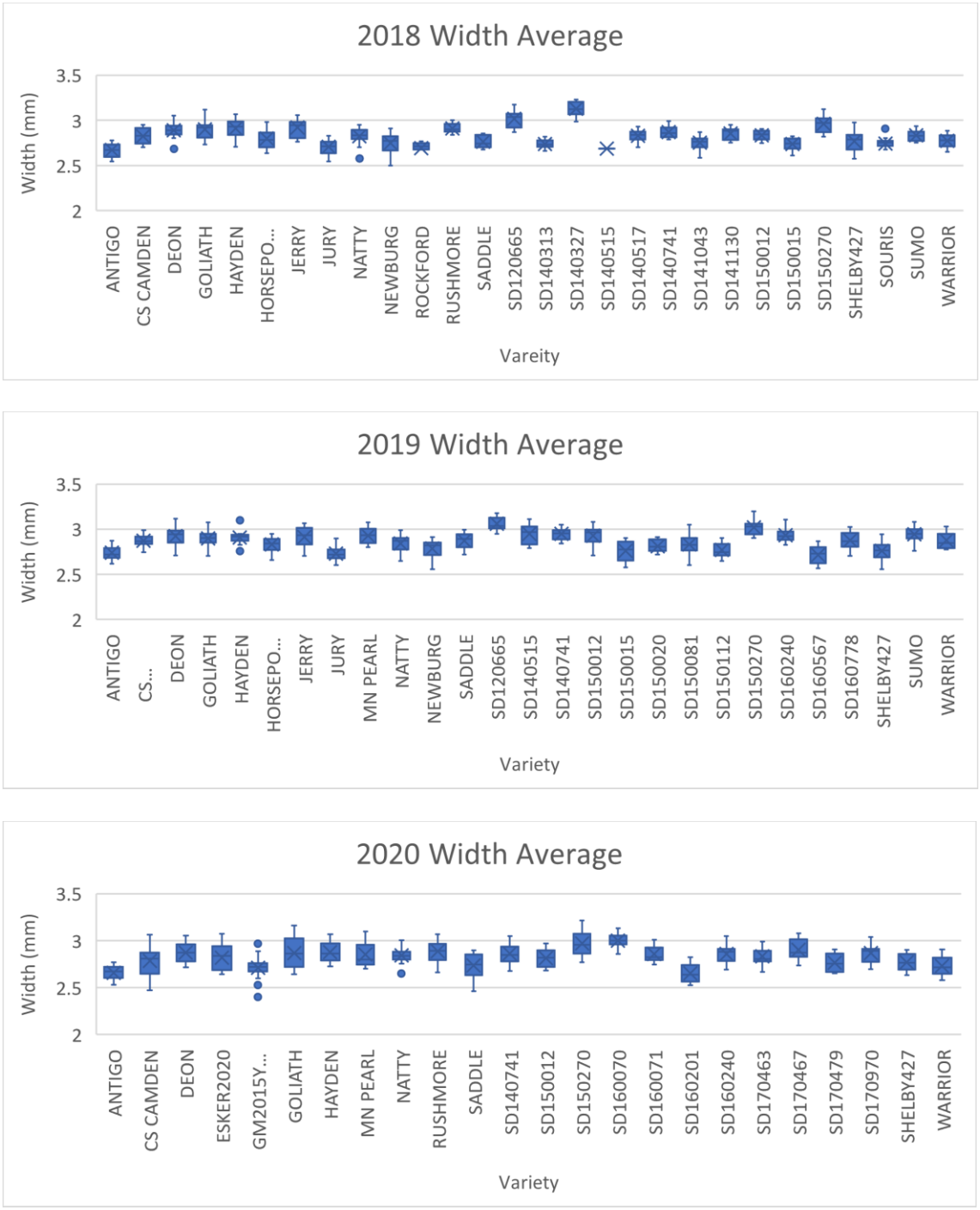
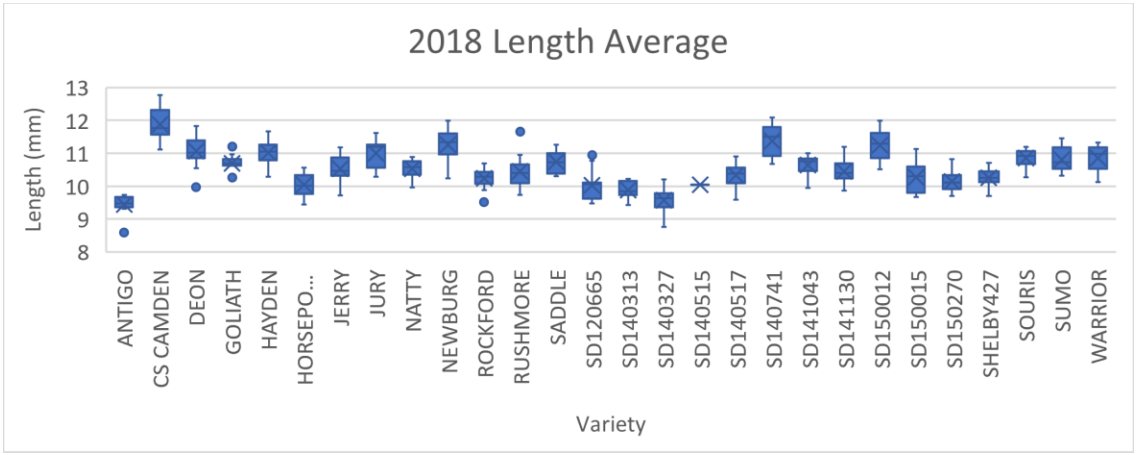
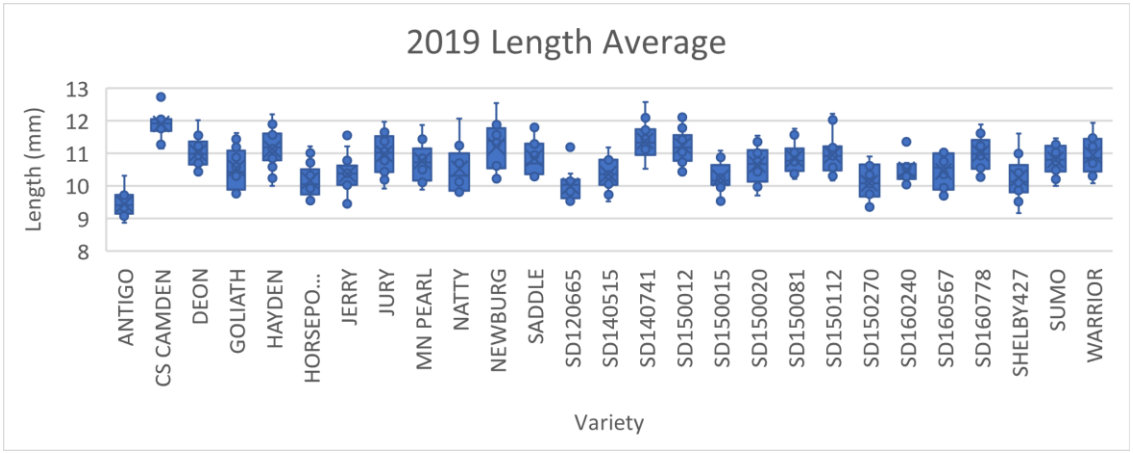


Figure 3. Box and whisker plots of oat kernels width (mm) for 27, 27 and 24 oat varieties evaluated at multiple locations in SD in 2018, 2019, and 2020.





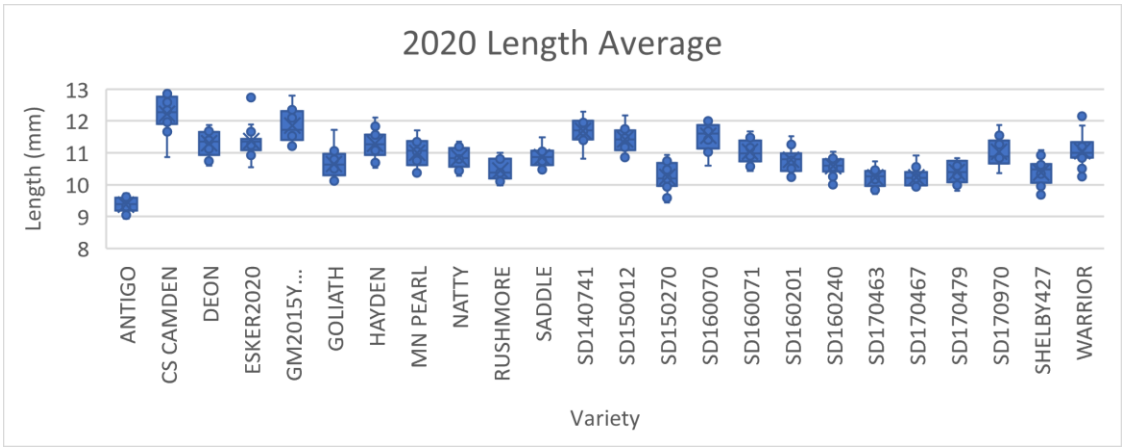
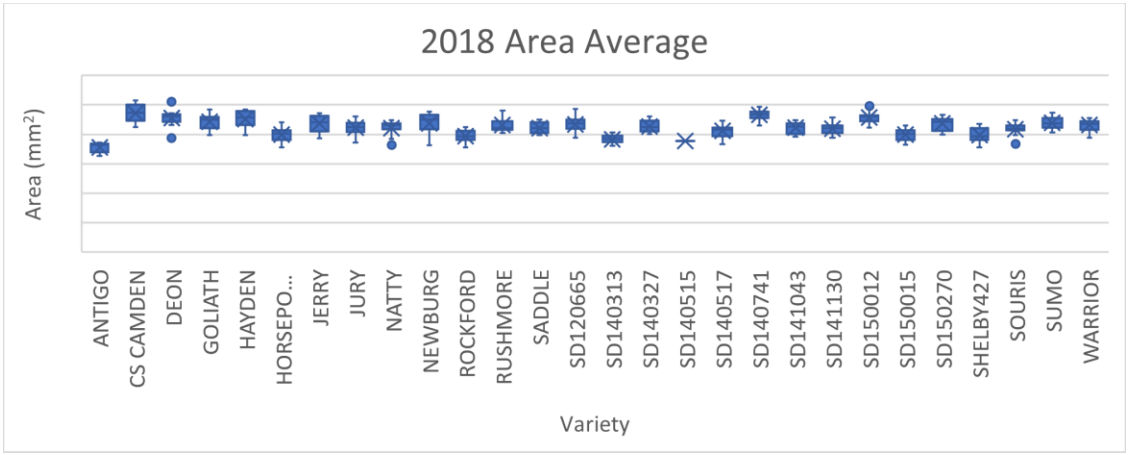
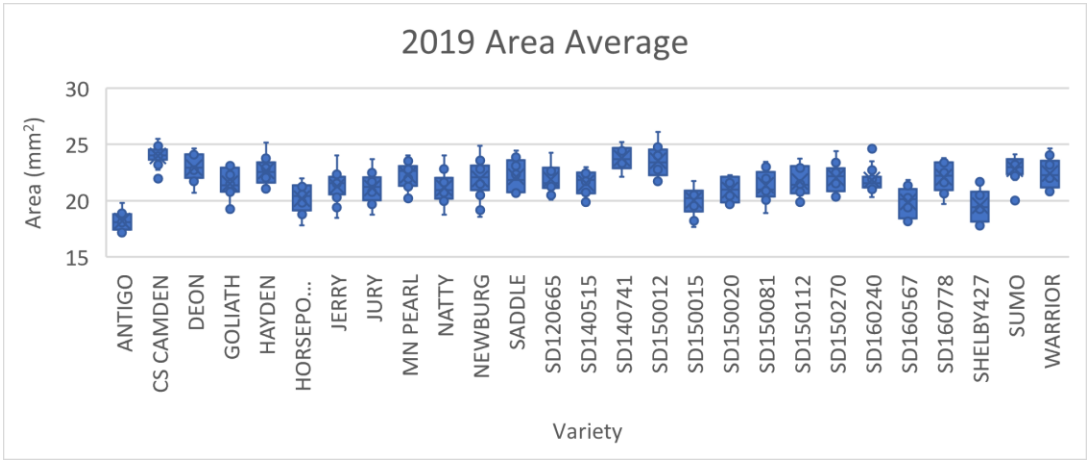


Figure 4. Box and whisker plots of oat kernel length (mm) for 27, 27 and 24 oat varieties evaluated at multiple locations in SD in 2018, 2019, and 2020.





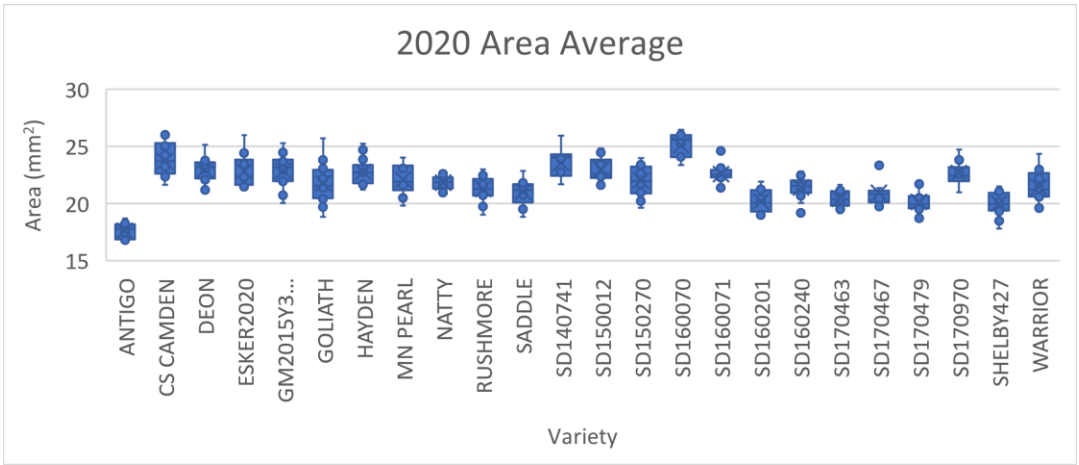


Figure 5. Box and whisker plots of oat kernel area (mm^2) for 27, 27 and 24 oat varieties evaluated at multiple locations in SD in 2018, 2019, and 2020.

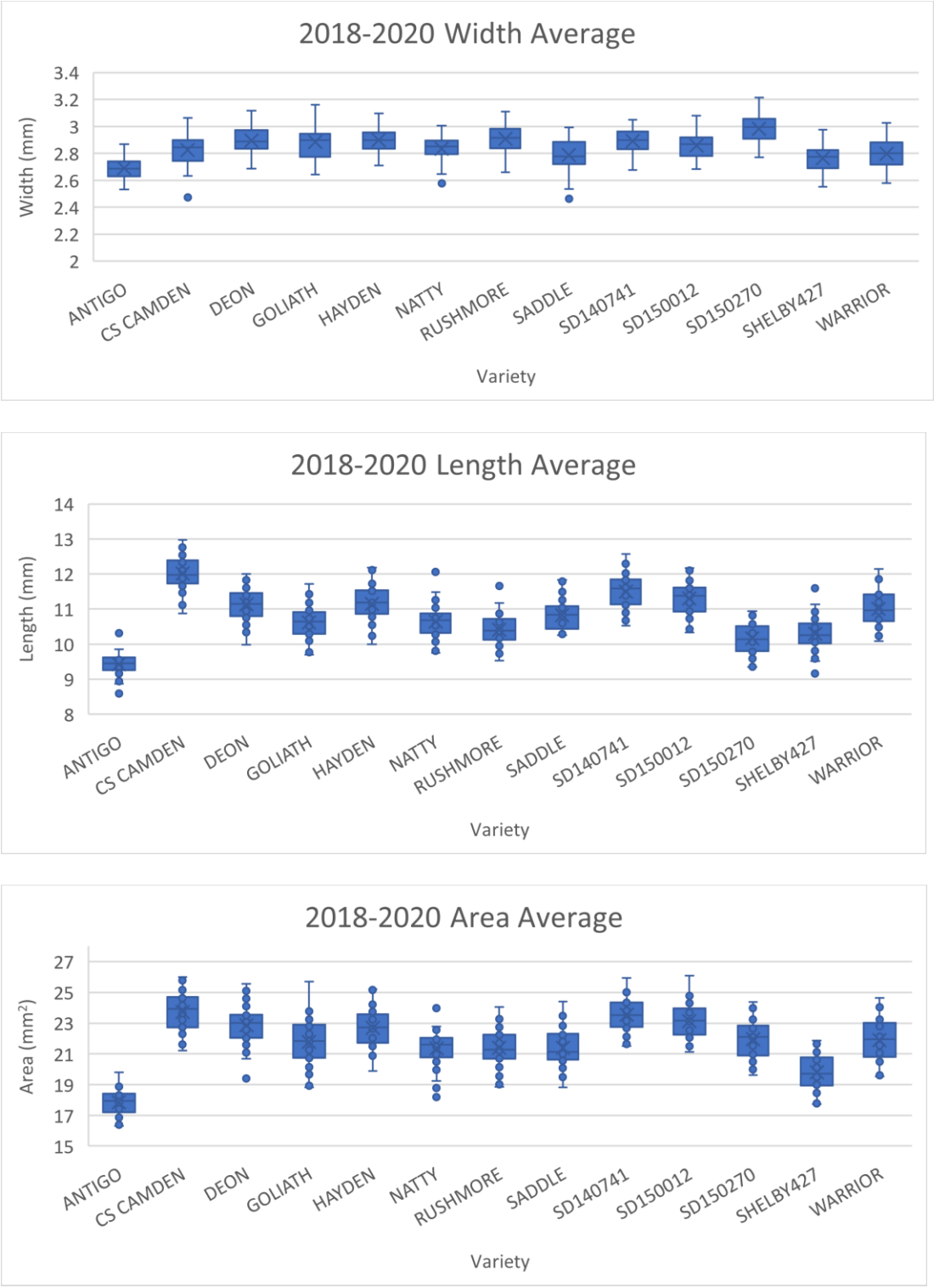


Figure 6. Box plot for width, length and area for 13 varieties that were common in all three years, evaluated at 21 locations from 2018 to 2020.

Table 3. Average kernel width for oat varieties evaluated in the SD Crop Performance Testing (CPT) from 2018 to 2020.

Average kernel width (mm)				
Genotype	2018	2019	2020	2018-2020
Antigo	2.67	2.73	2.66	2.69
CS Camden	2.82	2.87	2.78	2.82
Deon	2.89	2.92	2.87	2.89
Esker 2020			2.81	2
GM2015Y3232			2.71	
Goliath	2.90	2.89	2.86	2.88
Hayden	2.91	2.90	2.87	2.90
Horsepower	2.80	2.82		
Jerry	2.92	2.91		
Jury	2.70	2.73		
MN Pearl		2.92	2.85	
Natty	2.82	2.83	2.84	2.83
Newburg	2.75	2.78		
Rockford	2.69			
Rushmore	2.89	2.95	2.88	2.91
Saddle	2.76	2.87	2.73	2.79
Shelby 427	2.77	2.76	2.77	2.76
Souris	2.74			
Sumo	2.83	2.94		
Warrior	2.77	2.87	2.73	2.80
SD120665	3.01	3.06		
SD140313	2.74			
SD140327	3.12			
SD140517	2.83			
SD140741	2.87	2.94	2.86	2.89
SD141043	2.75			
SD141130	2.85			
SD150012	2.83	2.92	2.82	2.86
SD150015	2.74			
SD150270	2.96	3.01	2.97	2.98
SD150020		2.81		
SD150081		2.83		
SD150112		2.77		
SD160070			3.00	
SD160071			2.86	
SD160201			2.65	
SD160240		2.93	2.86	
SD160567		2.72		

SD160778		2.88		
SD170463			2.83	
SD170467			2.90	
SD170479			2.77	
SD170970			2.87	
Average	**2.83**	**2.87**	**2.82**	**2.85**
STD Average	**0.36**	**0.37**	**0.36**	**0.37**
LSD score	**0.061**	**0.064**	**0.047**	**0.034**

Table 4. Average kernel length for oat varieties evaluated in the SD Crop Performance Testing (CPT) from 2018 to 2020.

Average kernel length (mm)				
Genotype	2018	2019	2020	2018-2020
Antigo	9.44	9.44	9.38	9.42
CS Camden	11.83	11.90	12.22	12.02
Deon	11.06	11.01	11.30	11.13
Esker 2020			11.36	
GM2015Y3232			11.85	
Goliath	10.68	10.53	10.64	10.62
Hayden	11.01	11.13	11.29	11.15
Horsepower	10.04	10.15		
Jerry	10.53	10.35		
Jurry	11.00	10.98		
MN Perl		10.76	10.99	
Natty	10.52	10.51	10.86	10.65
Newberg	11.25	11.22		
Rockford	10.23			
Rushmore	10.44	10.36	10.45	10.40
Saddle	10.73	10.85	10.86	10.82
Shelby 427	10.25	10.24	10.38	10.29
Souris	10.85			
Sumo	10.81	10.80		
Warrior	10.85	10.94	11.06	11.01
SD120665	10.03	10.00		
SD140313	9.88			
SD140327	9.57			
SD140517	10.34			
SD140741	11.38	11.37	11.69	11.50
SD141043	10.64			
SD141130	10.45			
SD150012	11.24	11.22	11.42	11.30
SD150015	10.26	10.25		
SD150270	10.14	10.08	10.24	10.16
SD150020		10.65		
SD150081		10.84		
SD150112		10.94		
SD160070			11.54	
SD160071			11.03	
SD160201			10.78	
SD160240		10.49	10.57	
SD160567		10.43		

SD160778		11.00		
SD170463			10.22	
SD170467			10.23	
SD170479			10.40	
SD170970			11.03	
Average	**10.57**	**10.69**	**10.91**	**10.80**
STD Average	**1.61**	**1.62**	**1.72**	**1.67**
LSD score	**0.289**	**0.279**	**0.238**	**0.151**

Table 5. Average kernel area for oat varieties evaluated in the SD Crop Performance Testing (CPT) from 2018 to 2020.

Average kernel area (mm^2)				
Gneotype	2018	2019	2020	2018-2020
Antigo	17.76	18.17	17.62	17.85
CS Camden	23.57	23.96	23.66	23.74
Deon	22.79	22.91	22.92	22.88
Esker2020			22.78	
GM2015Y3232			22.84	
Goliath	22.12	21.50	21.77	21.77
Hayden	22.62	22.64	22.76	22.68
Horsepower	19.95	20.18		
Jerry	21.82	21.26		
Jurry	21.14	21.13		
MN Perl		22.24	22.08	
Natty	21.09	21.10	21.80	21.36
Newberg	21.91	21.93		
Rockford	19.65			
Rushmore	21.36	21.56	21.26	21.39
Saddle	21.07	22.13	21.00	21.40
Shelby427	19.85	19.54	19.98	19.79
Souris	20.91			
Sumo	21.99	22.82		
Warrior	21.55	22.40	21.44	21.90
SD120665	21.72	21.88		
SD140313	19.18			
SD140327	21.37			
SD140517	20.56			
SD140741	23.27	23.79	23.55	23.55
SD141043	21.10			
SD141130	20.97			
SD150012	22.88	23.51	22.98	23.12
SD150015	19.93	20.01		
SD150270	21.83	22.00	21.98	21.94
SD150020		20.95		
SD150081		21.32		
SD150112		21.68		
SD160070			25.09	
SD160071			22.56	
SD160201			20.18	
SD160240		21.83	21.41	
SD160567		19.88		

SD160778		22.07		
SD170463			20.40	
SD170467			20.93	
SD170479			20.11	
SD170970			22.64	
Average	**21.25**	**21.65**	**21.83**	**21.79**
STD Average	**5.39**	**5.48**	**5.78**	**5.66**
LSD score	**0.87**	**0.92**	**0.79**	**0.506**

Table 6. Average kernel width for oat genotypes evaluated at multiple locations in SD over three years.

Width (mm)			
Location	2018	2019	2020
Aberdeen	2.83	2.92	2.88
Dakota Lakes		2.88	2.73
Miller	2.86	2.86	2.98
NERF	2.87	2.95	2.92
SERF		2.79	2.71
Selby	2.91		2.82
Volga	2.73	2.83	2.80
Winner	2.76	2.84	2.74

Table 7. Average kernel length for oat genotypes evaluated at multiple locations in SD over three years.

Length (mm)			
Location	2018	2019	2020
Aberdeen	10.86	11.19	11.30
Dakota Lakes		10.59	10.50
Miller	10.84	10.64	10.88
NERF	10.58	10.67	11.02
SERF		10.18	10.79
Selby	10.31		10.45
Volga	10.15	11.48	11.11
Winner	10.68	10.10	11.21

Table 8. Average kernel area for oat genotypes evaluated at multiple locations in SD over three years.

Area (mm^2)			
Genotype	2018	2019	2020
Aberdeen	22.00	22.74	22.65
Dakota Lakes		21.97	20.61
Miller	21.91	20.86	22.62
NERF	21.64	21.89	22.91
SERF		20.36	21.19
Selby	21.66		21.25
Volga	19.70	22.41	21.81
Winner	20.56	20.68	21.52

Table 9. The relative proportion of variance components associated with variation in kernel width, length, and area for oat genotypes evaluated in the South Dakota Crop Performance Testing from 2018 to 2020.

		Relative proportion of variance (%)		
	Variance component	Width	Length	Area
2018	Environment	25.6%	17.9%	26.9%
	Genotype	57.5%	66.5%	52.8%
	Environment*Genotype	1.4%	2.6%	1.5%
	Residuals	15.5%	14.0%	18.8%
2019	Environment	21.2%	42.8%	22.4%
	Genotype	46.0%	42.8%	50.1%
	Environment/Genotype	9.6%	2.0%	4.2%
	Error	23.1%	12.2%	23.1%
2020	Environment	46.2%	17.7%	18.0%
	Genotype	35.5%	70.3%	62.2%
	Environment/Genotype	6.6%	1.7%	2.8%
	Error	11.7%	10.3%	17.0%
2018-2020	Environment	38.1%	21.4%	16.8%
	Genotype	33.6%	67.1%	62.7%
	Environment/Genotype	9.0%	2.0%	3.4%
	Error	19.2%	9.3%	17.0

Table 10. Broad-sense heritability estimates for kernel width, length and area for oat genotypes evaluated in the SD Crop performance Testing from 2018 to 2020.

Broad-sense heritability estimate			
Year	Width	Length	Area
2018	0.69	0.79	0.66
2019	0.68	0.5	0.69
2020	0.43	0.79	0.78
2018-2020	0.46	0.76	0.79

Table 11. Pearson correlation coefficients between kernel size (width, length and area) and key milling quality characteristics.

Pearson correlation with width		
	Correlation coefficient	P-value
Protein content	-0.0851	0.07189
Beta glucan content	-0.1373	0.0035**
Oil content	-0.1332	0.0047**
Test weight	0.1068	0.0289*
STDEV width	0.1925	0.0000***
Groat percent	0.3963	0.0000***
Thousand kernel weight	0.6067	0.0000***
Plump	0.6786	0.0000***
Mid	-0.6459	0.0000***
Thin	-0.5218	0.0000***
Pearson correlation with length correlation		
Protein content	-0.2752	0.0000***
Beta glucan content	0.2375	0.0000***
Oil content	-0.2052	0.0000***
Test weight	-0.4985	0.0000***
STDEV Length	0.0775	0.0742
Groat Percent	-0.2009	0.0000***
Thousand kernel weight	0.3789	0.0000***
Plump	0.1107	0.0191*
Mid	-0.0742	0.1169
Thin	-0.1936	0.0000***
Area Correlation		
Protein content	-0.1831	0.0000***
Beta glucan content	0.1048	0.0265*
Oil content	-0.2176	0.0000***
Test weight	-0.2430	0.0000***
STDEV Length	0.1670	0.0557
Groat Percent	0.0820	0.0833
Thousand kernel weight	0.6338	0.0000***
Plump	0.4509	0.0000***
Mid	-0.4008	0.0000***
Thin	-0.4428	0.0000***

CHAPTER 2: GENOME-WIDE ASSOCIATION MAPPING FOR KERNEL SIZE IN OATS

ABSTRACT

Oat kernel size is a crucial milling characteristic because large kernels can make larger, more desirable oat flakes. In addition, it is easier for the milling industry to separate gluten contaminants when oat varieties have large kernels. A better understanding of the genetic control of kernel size would be helpful for oat breeders. To further research this issue, a genome-wide association mapping study was performed to identify QTLs (Quantitative trait loci) and markers associated with kernel width, length, and area in oats. The kernel width, length, and area were measured on grain samples from 234 oat breeding lines grown in 4 locations in South Dakota using a Qualmaster Computer Vision Device (VIBE) (Vibe Imaging Analytics Ltd., Ben-Brak, Israel). Genotyping was completed using genotyping-by-sequencing. A total of 7,979 SNP markers were used for the genome-wide association study (GWAS) to identify the locations of significant QTLs/markers that determine kernel size in oats. In this study, four unique markers showed significant association with kernel size, three were significantly associated with kernel width, and one was significantly associated with kernel length. Using the newly published oat genome, we identified the RNA sequences closest to these markers. NCBI blast was used to see if a similar sequence was present in the genome of another species. Of the four unique markers, only one had an RNA sequence that was found in another species. That sequence was found in Emmer; the function of the locus has not been determined yet; further research needs to be done to determine the function of that locus.

2.1 INTRODUCTION

Kernel size (width, length, and area) is becoming an essential trait in oats (*Avena sativa* L.), making breeding larger kernel size more critical. Milling facilities desire varieties with larger oat kernels because contaminant seeds (wheat, barley, and rye) are easier to separate when oats have large kernels. Smaller oat kernels fall into the category of "contaminant seed" in milling facilities and results in waste. For these reasons, milling facilities have encouraged breeders to breed for oats with larger kernels (width, length, and area). It is very beneficial for breeders to understand the genetic control of grain size to make sure the new varieties produce large kernels and meet the milling facility standards.

We used genome wide association study (GWAS) to gain some understanding on the genetic control of kernel size in oats. GWAS requires the use of a diverse population with genetic variation for a particular trait of interest. Our traits of interest are width, length and area. GWAS scans the genome looking for markers that can be used to predict the presence of the trait of interest. There has not been a study for GAWS looking at kernel size in oats, but there have been other traits investigated, including lodging resistance, fusarium head blight resistance, fatty acid composition, and seed vigor. A GWAS study looking at lodging found six makers significantly associated with lodging resistance (Tumino, 2017). One study investigated fusarium head blight, did not find any significant markers; however, they did find some resistant-related traits that showed cross-validation accuracy when making crosses (Hikka, 2020). A study looking at ten fatty acid composition found 148 significant single nucleotide polymorphisms (SNP)

(Carlson, 2019). Seed vigor has also been evaluated. The study found 36 and 16 unique loci associated with root and shoot traits (Huang, 2020).

The genetic control of grain size has been evaluated in other cereal crops. In rice, two genes, qSW5/GW5 and GW2, determine seed width and weight (Song Yan, 2011). They found that when GW2 expression was suppressed, the seed length, thickness, and weight significantly increased (Song Yan, 2011). In wheat, thousand kernel weight (TKW) has been linked to loci on chromosome 6D and 4A for high TKW and 5B and 5D for low TKW (J.S.S. Ammiraju, 2001). According to Flavio Breseghello, 2006, the markers Xumc111 and Xgwm30 showed significant association with kernel width in wheat (Flavio Breseghello, 2006). Another study in wheat found that chromosome 4B contributed to 40% of grain dimension variation (Moran Nave R. A.-Z., 2016).

The reference genome for oats was published in the spring of 2020 (https://wheat.pw.usda.gov/jb/?data=%2Fggds%2Foat-ot3098-pepsico&loc=5C%3A50997567..51001131&tracks=DNA%2Cpepsi%2Chu%2C6k-probes&highlight=). Having a reference genome, we now have the ability to know what chromosomes to link the markers to and can determine how close significant markers are to one another.

For this project, our goal was to identify significant quantitative trait loci (QTL) that determine kernel size, more specifically, length, width, and area in oats. We want to know their location so that we can select varieties with the molecular markers associated with the QTLs for larger kernel size.

2.2 Materials and Methods

2.2.1 Phenotypic data collection

Grain samples from the 2019 preliminary yield trial (PYT) from the South Dakota State University oat breeding program were used for this study. A total of 234 breeding lines were evaluated along with four replicated checks at Winner, the Southeast Research Farm (SERF) in Beresford, the Northeast Research Farm (NERF) in South Shore, and Volga. All locations are in South Dakota. The field layout followed an augmented design in all locations.

Kernel size (length, width, and area) was measured using a VIBE (Vibe Imaging Analytics Ltd., Ben-Brak, Israel). The methodology used is described in Chapter 1, section 1.2.2.

2.2.2 Phenotypic data analysis

There were histograms made in R (R Core Team 2020) to look at the phenotypic data and distinguish outliers.

Best Linear Unbiased Predictors were computed using a linear mixed model with the minque package in R (Wu, 2019). The broad-sense heritability for each trait was calculated using the same methods as in Chapter 1. Refer to section 1.2.4.

2.2.3 Genotyping

The tissue samples for the DNA extraction were grown in 72 well seed starting trays for ten days. For each genotype, six wells planted were with two seeds per well for a total of twelve plants per genotype. This was done to ensure there was enough tissue to be used for DNA extraction. On Day 10, the tissue was harvested, the top half-inch of the leaves of 6 plants were cut off, and the corresponding leaf tissue samples were placed in a 96 well plate. An air-pore sheet was placed over the samples, so the tissue can still

breathe and let excess moisture escape. Silica sand was placed in the plastic bag with the 96 well plate to absorb the moisture from the tissue samples.

The samples were sent overnight to HPI (Manhattan, KS) for DNA extraction and sequencing. DNA was extracted via the Qiagen DNeasy 96 Plant Kit following the manufacturer's instructions. Genotype-by-sequencing (GBS) libraries were generated via Mspl-Pstl double digest and barcoded using 384 unique barcodes ligation before pooling the markers. Sequencing was performed via Illumina NexSeq 550 to generate single-end 100bp reads. To call the single nucleotide polymorphism (SNP), the fastq files are trimmed for base quality and the removal of Illumina adapter readthrough using trim_galore (https://github.com/FelixKrueger/TrimGalore) with additional filtering requiring all of the reads to be at least 74 bp long. The SNPs were called using Tassel 5 (Bardbury, 2007) ProductionSNPCallerPluginV2 against a defined set of 10905 oat markers (built from hundreds of oat samples and filtered with a minor-allele-frequency of at least 2%) anchored against the OT3098 v1 reference oat genome. The SNPs were then filtered to have maximum heterozygosity of no larger than 50%.

After initial genotyping, minor alleles with a frequency of 0.02 or less were removed. Missing data were imputed with 'A.mat' in the R package rrBLUP (Endelman, 2019). Markers with data for less than 10% of the genotypes were removed. Also, markers with more than 50% of heterozygous calls were removed. With these outlines, any markers with less than 50% calls across the sample were removed. This resulted in 7,979 markers for analysis. Once the SNP reads were cleaned up, the R package rrBLUP was used to run the GWAS.

2.2.4 Association analysis

To analyze the results for association mapping, the R package rrBLUP (Endelman, 2019). The model used is $y = x\beta + Zg + S_T + \backslash varepsilon$ where β is a vector of fixed effects modeling both environmental factors and population structure. G is the genetic background of each line as random effect. The variable T models the additive SNP effect as a fixed effect. The residual variance is the varepsilon. Kinship was not taken into account in our model, because ours was a diverse population according to our principal component analysis. We used a minimum allele frequency of 0.05. There were 1,077 phenotypic observations used and 7,909 SNP markers used for our GWAS analysis. The R package ggplot2 (de Vries 2020) was used to make the principal component analysis. Tassel 5 (Bardbury, 2007) was used to make the linkage disequilibrium graph. The graph is compiled of the r^2 value and the base pair distance.

2.3 Results

2.3.1 Phenotypic measurement and heritability

Histograms were made to evaluate the frequency distribution for each kernel size measurement (width, length, and area). The histograms are in Figure 7 for width, Figure 8 for length, and Figure 9 for area. It was determined that the threshold for width would be between 2.25 mm – 3.5 mm, and between 8 mm and 13 mm for length. With these thresholds, some data points were discarded because they were outliers.

Kernel Width

There needs to be variation in kernel size for the different varieties in our association mapping population for association mapping to work. The smallest kernel width observed was 2.34 mm at both SERF and Volga for experimental lines SD180139 and SD180262, respectively (Table 12). These two locations have heavy disease pressure,

which can contribute to the production of small kernels. The maximum kernel width (2.87mm) was observed at the NERF for experimental line SD180770 (Table12). On average, kernels with smaller widths (2.7mm) were produced in Winner, while kernels with larger width (2.87mm) were grown at the NERF. The broad-sense heritability estimate for kernel width was 0.85 (Table 13). This heritability is significantly higher than the heritability estimates measured in the SD CPT oat variety trials.

Kernel Length

Among the oat grain samples from 238 genotypes grown at four locations in SD, the sample with the shortest length (8.5 mm) was for SD141192 (Table 12) grown at Winner, and the sample with the longest kernel length (12.77mm) was for SD180276 grown at Volga. On average, longer kernels were produced in Winner (11.49 mm), while shorter kernels were produced at the NERF (9.83 mm on average) (Table 12). The broad-sense heritability estimate was 0.47 (Table 13), which is lower than what we observed in the set of genotypes evaluated in the SD CPT oat variety trials.

Kernel Area

The largest average kernel area was measured at Volga (27.47 mm^2) for SD180266, and the smallest average kernel area was measured at SERF (16.88 mm^2) for experimental line SD180341 (Table 13). Kernels with the largest area (21.10 mm^2 on average) were produced at the SERF, while kernels with the smaller area (20.09 mm^2) were produced in Winner (Table 13). This result is an interesting outcome because the SERF has a lot of disease pressure (crown rust), leading to the production of kernels with smaller area. The broad-sense heritability estimate for kernel area was 0.88 (Table 13).

This number is higher than what we saw in the set of genotypes evaluated in the SD CPT oat variety trial.

2.3.2 Population Structure and Linkage Disequilibrium

The marker classification was graphed to get a visual representation of the reads per variety. The results of this classification are in Figure 10. From this marker classification, you can see there were good DNA reads for our genotypic data. Only one genotype did not result in good reads, and it was discarded from the analysis.

Relatedness was also evaluated and principal component analysis (PCA) was conducted. With association mapping, the mapping population should be diverse. While looking at our PCA in Figure 11, some clusters are observable, but we have a diverse population. Since the population is from a breeding program and the PYT specifically, some sibling experimental lines were tested. The PCA is composed of PC1 and PC2 components. PC1 and PC2 have a cumulative proportion of 0.09 and 0.13. The standard deviation is 13.06 and 9.09 for PC1 and PC2, respectively.

A linkage disequilibrium analysis was performed for our population. We found that at approximately $r^2 = 0.20$, there is no more linkage disequilibrium (Figure 12).

2.3.3 Association analysis

Results for the GWAS and Manhattan plots are presented in Figures 13, 14, and 15, for area, length, and width, respectively. A Bonferroni correction was performed, and a significance level of 0.01 was used. The logarithm of odds (LOD) score was 5.902. There were eight marker/location combinations with significant LOD scores, but only four unique markers were significant. The markers and their chromosome location are listed in Table 14. There was one marker significant for length and seven markers

significant for width, and there was no marker significant for area. All of the markers are located on chromosomes 5A and 5C. The markers with a significant LOD score are 51000364, 51399270 on chromosome 5C, and 137401828, 181475218 on chromosome 5A (Table14). The closest RNA sequence for the four unique markers was identified by referencing the reference genome in Grain Genes (Victoria Blake, 2019). It was calculated to identify how close the marker is to the most immediate RNA sequence (Table 15). Once the nearest RNA sequence was obtained, the sequence was taken to the NCBI blast website (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome) to determine if a similar sequence was present in the genome of another species. Of the four unique markers, only one (5C_51399270) was found in another species: Emmer, with a 90.48% homology. The sequence in Emmer was for an uncharacterized locus (LOC119354425). Since this is an uncharacterized locus, we do not know its function.

2.4 DISCUSSION

The oat breeding population used in this study showed a wide range in kernel size. The kernel size varied between 2.34 mm – 3.26 mm for width, 8.5 mm – 12.75 mm for length, and 16.76 mm – 27.47 mm for area. The mapping population used had a wider range for width, length, and area than our CPT population in the previous chapter. This shows that despite the use of a breeding population from a single breeding program, we did have a broad range of genotypes tested in our population.

The broad-sense heritability estimates in this population were high for kernel width and area and moderate for kernel length; this is different than what we saw in the

CPT oat variety trials. In our mapping population, we saw a much higher heritability for width, which coincides with the number of markers significantly correlated with width size. The heritability estimate for length was not as high (higher in the CPT oat variety trial), with only one significant marker being detected. The heritability estimates for area was high in both populations but there were no significant markers detected for area in our study, which was surprising.

The oat genome is not fully annotated yet, however some RNA sequences that are relatively close to the significant markers were identified, but there is still some distance between marker location and RNA sequence. There could be RNA sequences closer to the markers significantly associated with kernel size in our population, but not annotated yet that have been found to affect kernel size in another crop. In order to find RNA sequences that are closer to our markers, more time and research will need to occur for the oat genome to be more annotated. There is currently a pan oat genome project in process that is working on mapping 30 different oat varieties. With only a single oat variety being mapped, it is hard to know the exact genes that determine certain traits. Many crosses and inversions can happen to the genome. Due to containing three genomes in oats. Once that gets done, we will have a better understanding of the oat genome.

Since we know of some markers that affect kernel size, breeders can use marker assisted selection for those markers to breed for larger kernels in oats. It is also known that length, width and area are heritable traits and that genetic gains can be made through breeding efforts. There is little genotype by environment interaction, indicating that there will be little rank change between varieties. The largest kernelled varieties will always be

the largest from environment to environment. This helps breeders in using the largest kernelled varieties for parents, because they will pass the large kernels to their progeny.

Milling facilities benefit from this study by knowing that the largest varieties will always be the largest, no matter the environment. Since there is little genotype by environment interaction. The larger kernel size also avoids excess milling waste when it comes to separating out contaminants. Larger kernels also can produce larger flakes, whcih make better oatmeal and other food products. Larger flakes make the end product for the oat more versatile. In addition, the larger kernels also are easier to dehull, resulting in conserved energy in the dehulling process.

FIGURES AND TABLES

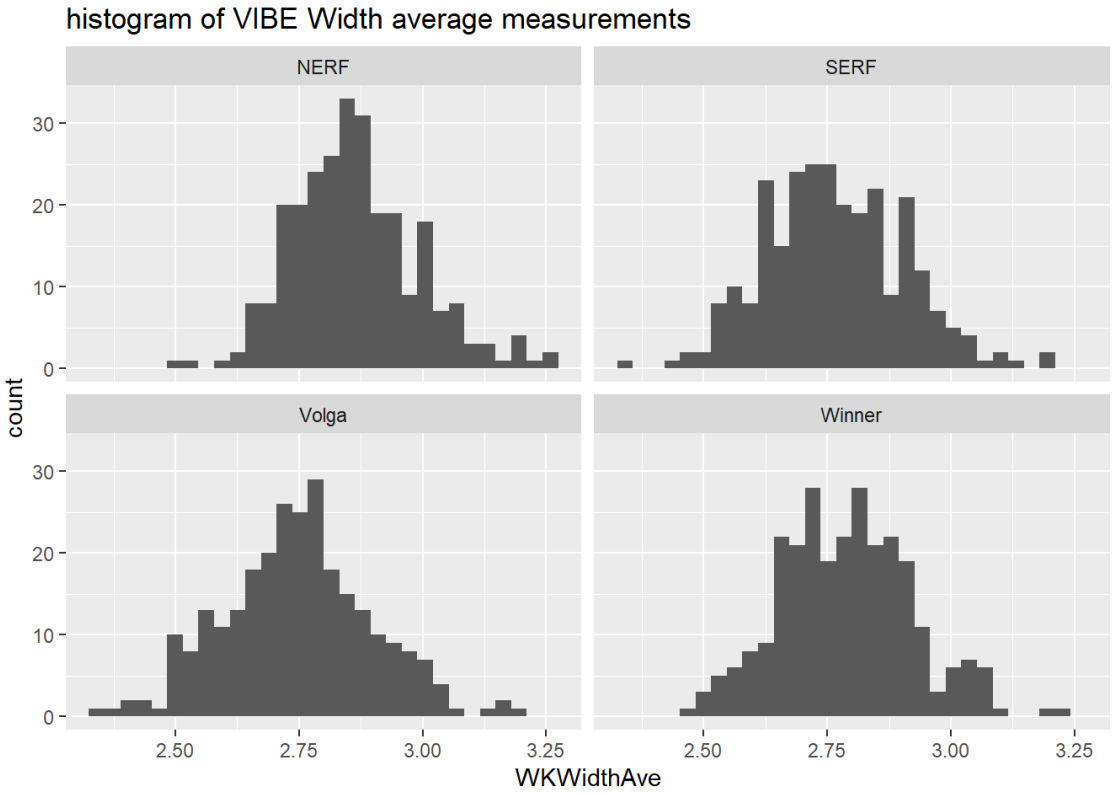


Figure 7. Frequency distribution of kernel width (mm) for grain samples from 238 oat genotypes grown at four locations in South Dakota.

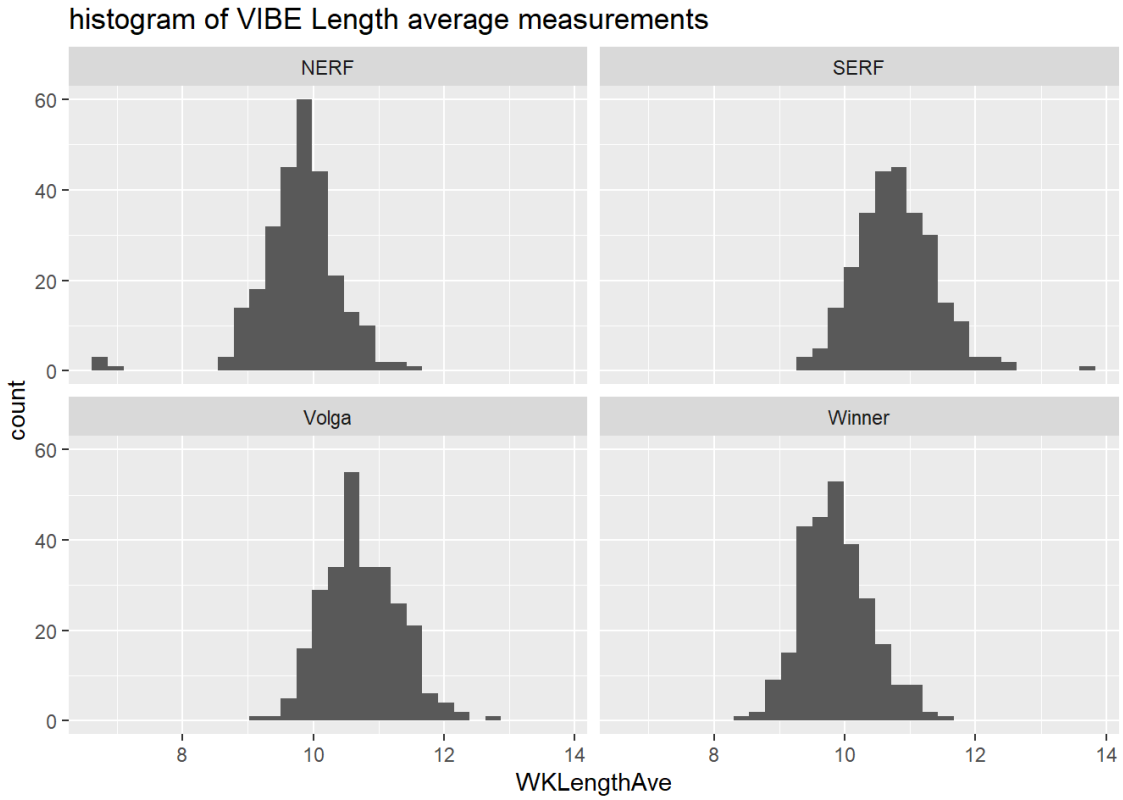


Figure 8. Frequency distribution of kernel length (mm) for grain samples from 238 oat genotypes grown at four South Dakota locations. A few outliers are visible at NERF and SERF. Those data points were discarded from our analysis.

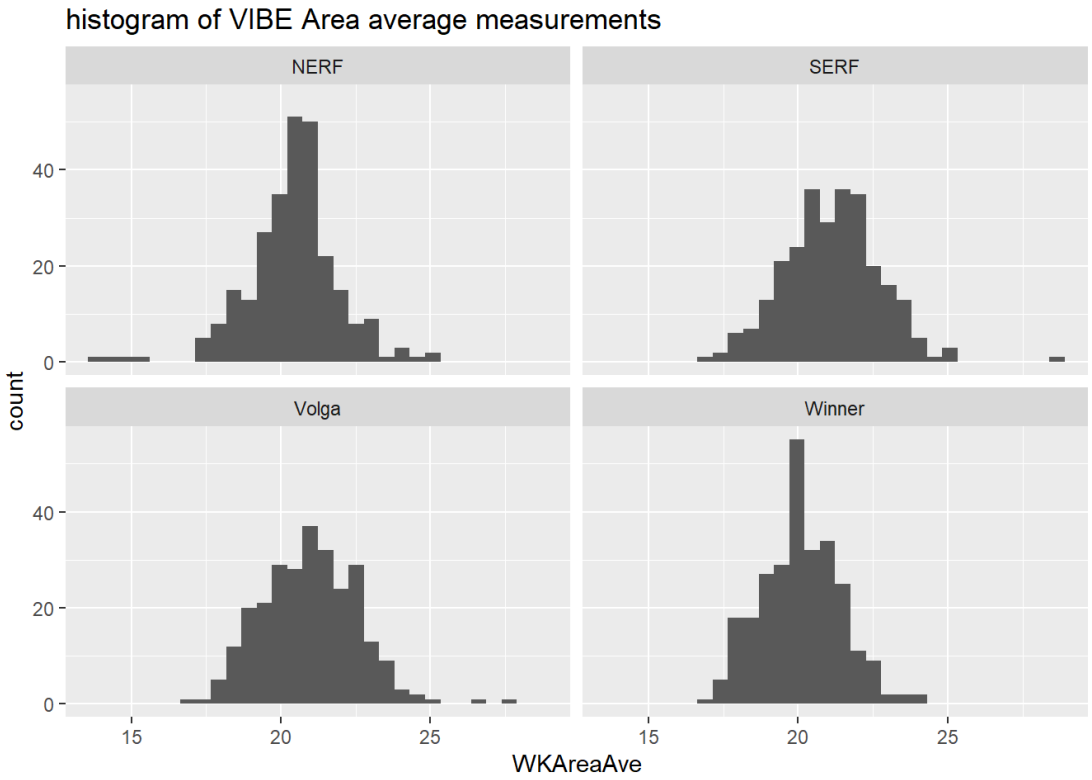


Figure 9. Frequency distribution of kernel area (mm^2) for grain samples from 238 oat genotypes grown at four South Dakota locations. A few outliers are visible at NERF and SERF. Those data points were discarded from our analysis.

Table 12. Minimum, maximum, and average kernel width, length, and area for 238 oat genotypes evaluated in four locations in South Dakota.

	Min	Variety	Max	Variety	Average
Environment	Width (mm)				
NERF	2.55	SD180026	3.26	SD180770	2.87
SERF	2.34	SD180139	3.19	SD180703	2.77
Volga	2.34	SD180262	3.18	SD180770	2.75
Winner	2.45	SD181172	3.24	SD180770	2.7
Environment	Length (mm)				
NERF	8.62	SD180420	11.48	SD180276	9.83
SERF	9.33	SD180369	12.45	SD181021	10.78
Volga	9.22	SD180703	12.77	SD180276	10.74
Winner	8.5	SD141192	11.49	SD180012	11.49
Environment	Area (mm^2)				
NERF	17.25	SD180238	25.23	SD180266	20.49
SERF	16.88	SD180341	25.31	SD180770	21.10
Volga	16.97	SD180238	27.47	SD180266	20.98
Winner	16.76	SD180239	24.20	SD180266	20.09

Table 13. Broad-sense heritability estimates for width, length, and area for a population of 238 oat genotypes grown at 4 locations in SD.

	Broad- Sense Heritability
Width	0.85
Length	0.47
Area	0.88

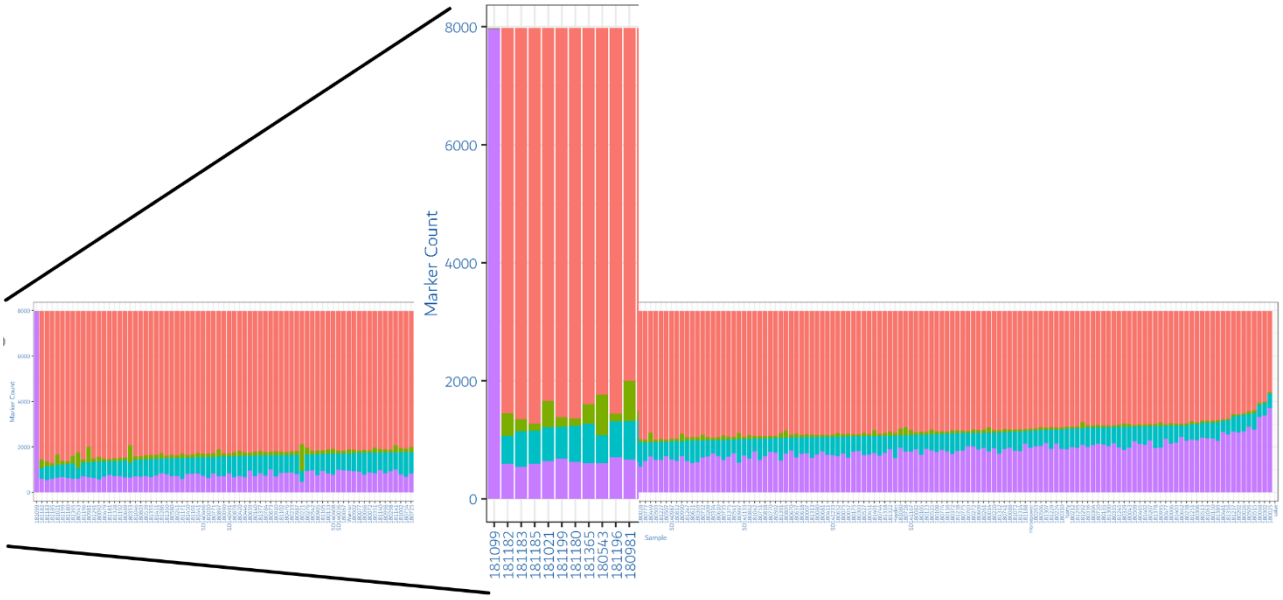


Figure 10. Marker classification following DNA sequencing of 238 oat genotypes using genotyping by sequencing. Orange is homozygous wild type, green is heterozygous, blue is homozygous for opposite of wild type and purple is no reads.

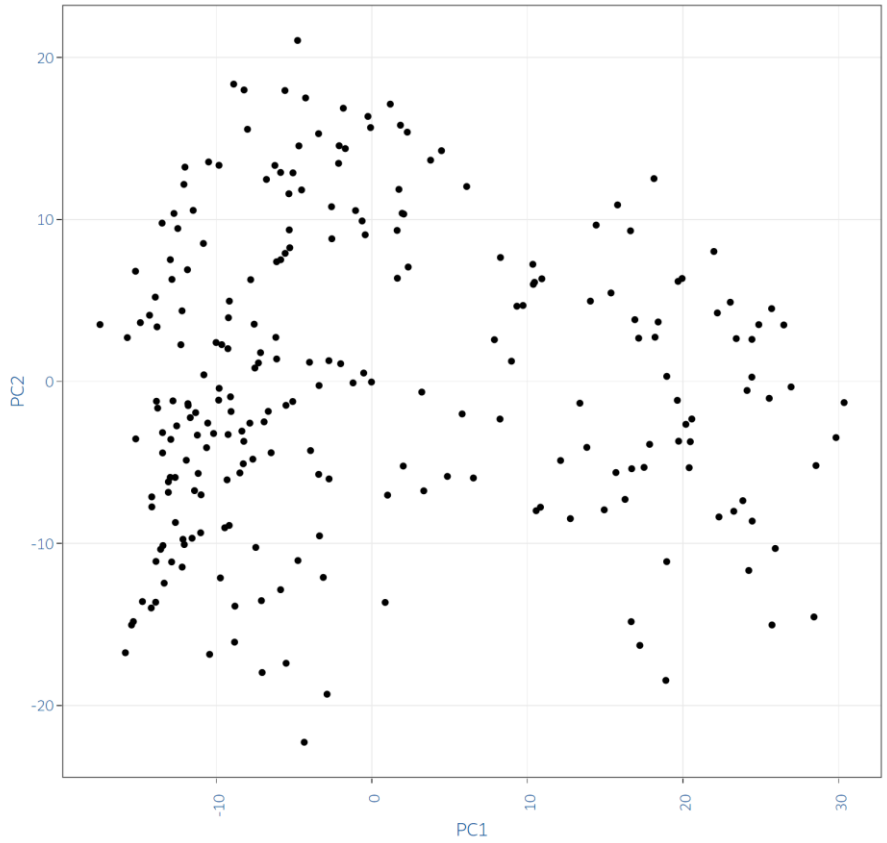


Figure 11. Principal component analysis (PCA) of 238 oat genotypes using genotypic data looking at the genetic diversity of the mapping population.

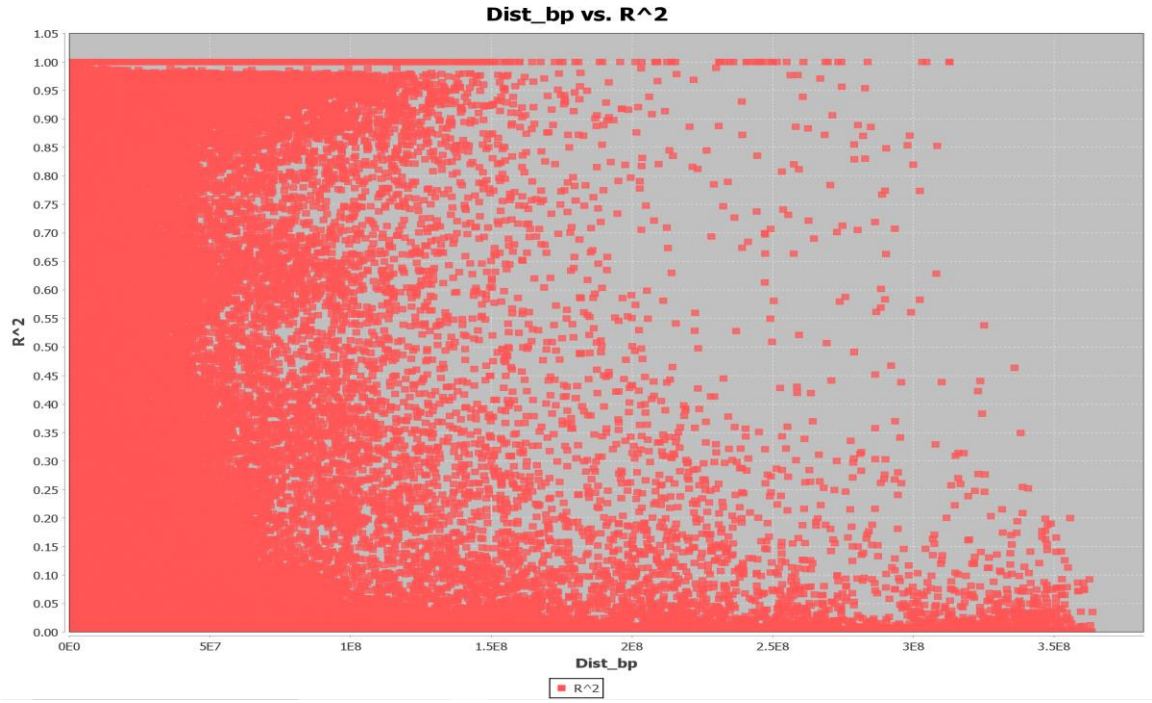


Figure 12. Linkage disequilibrium graph for oat genotypes included in the 2019 preliminary yield trial.

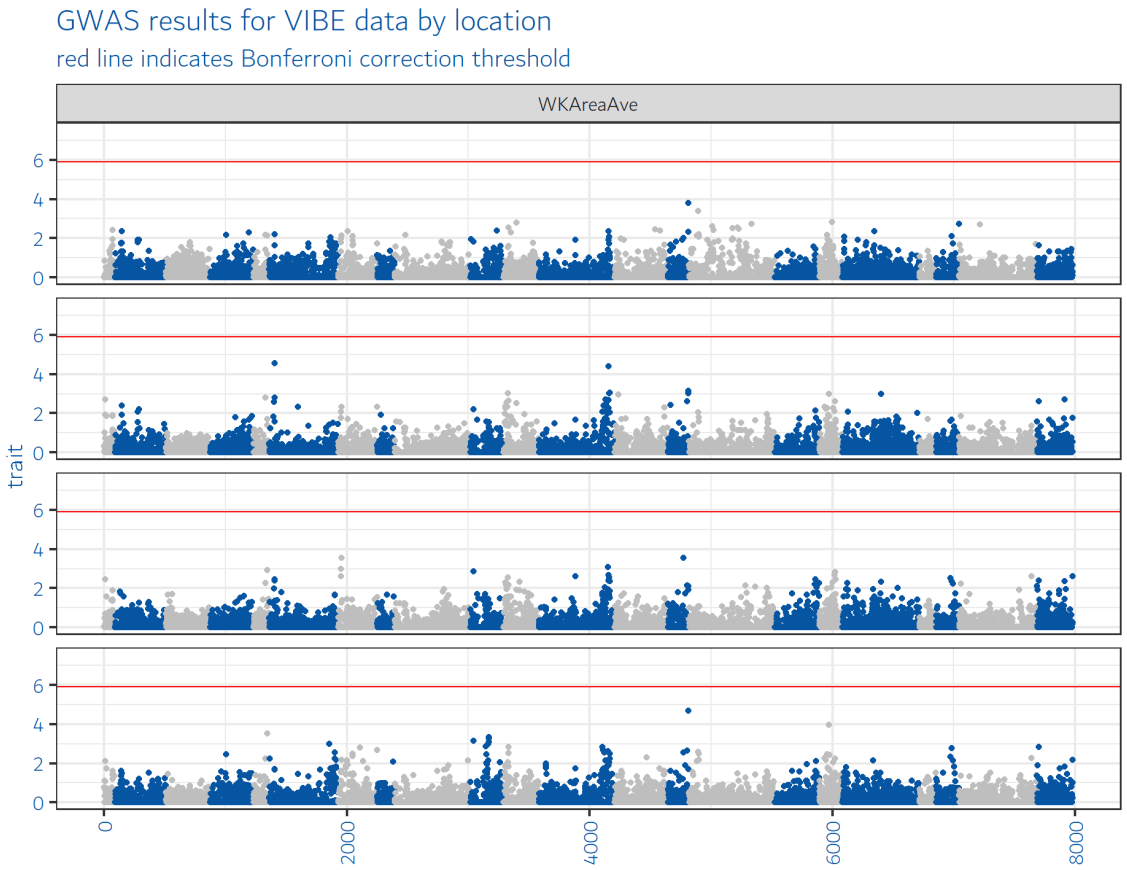


Figure 13. Manhattan plot for oat kernel area. There are not significant markers above the Bonferroni correction threshold.

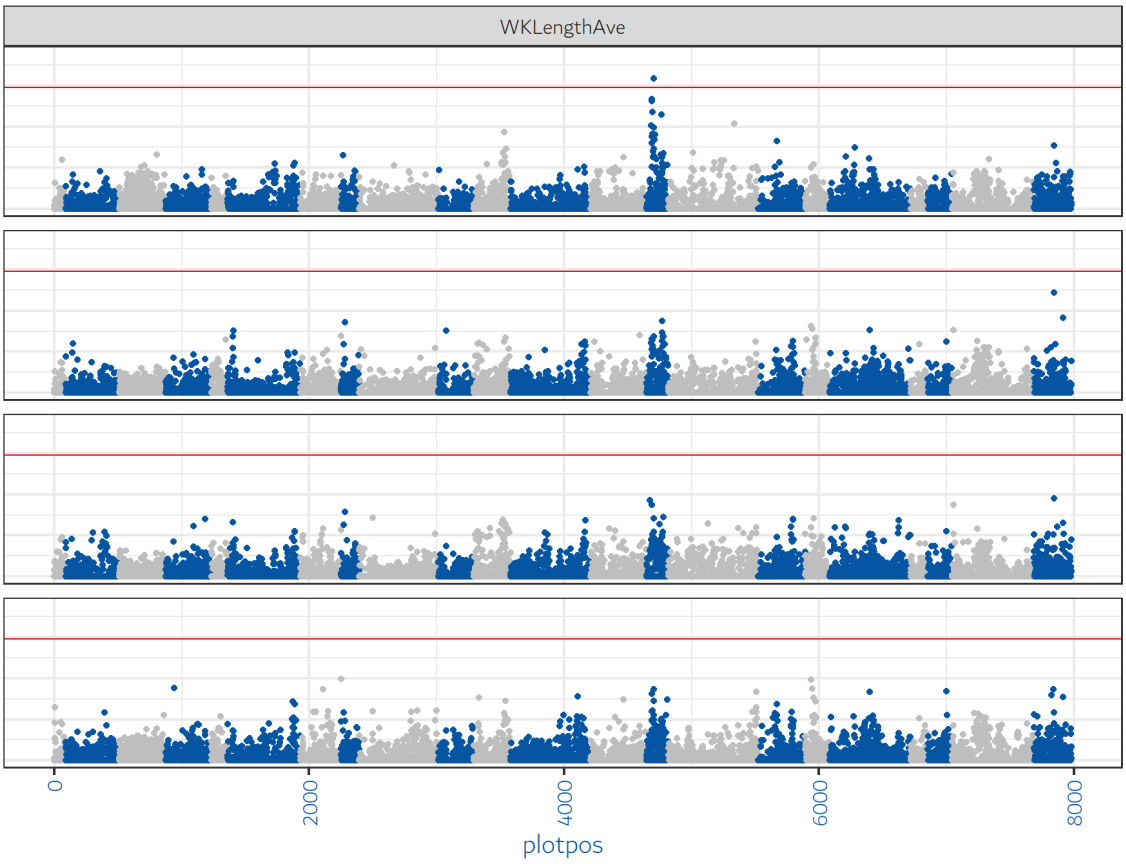


Figure 14. Manhattan plot for oat kernel length. There is one marker significant for length based on the Bonferroni correction.

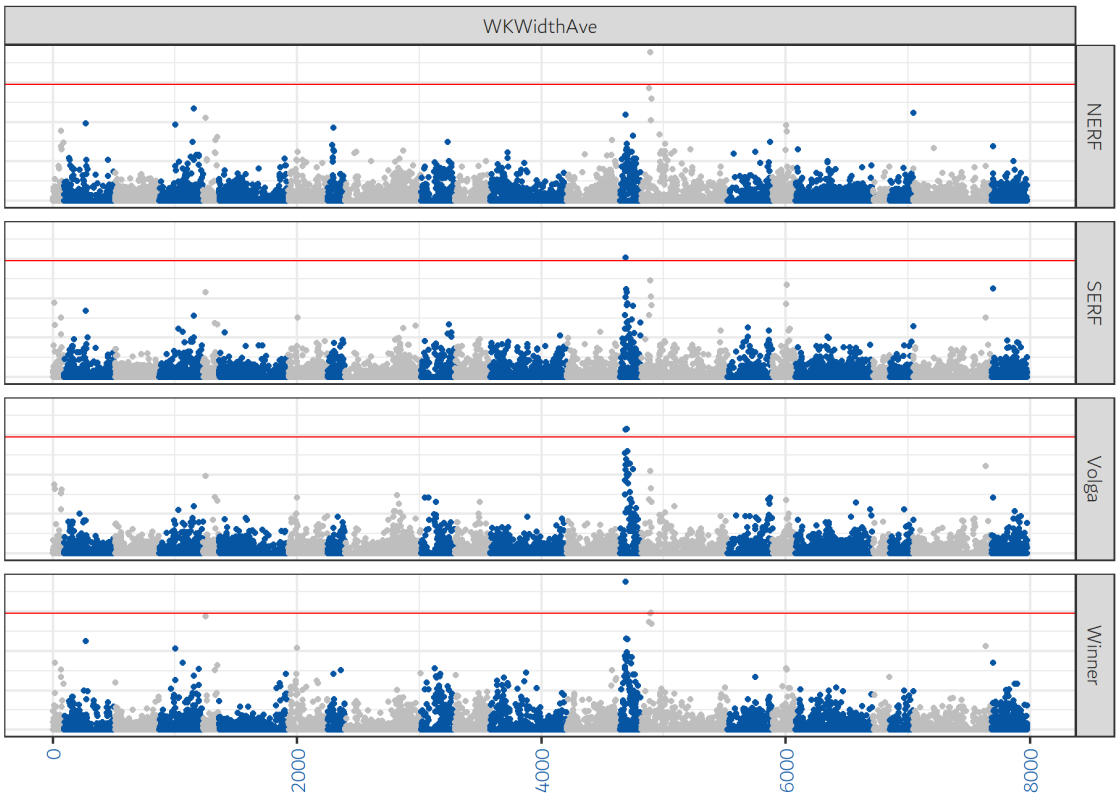


Figure 15. Manhattan plot for oat kernel width. There are multiple markers significant for width based on the Bonferroni correction.

Table 14. List of significant markers, their chromosome, LOD score, trait and location.

Marker Name	Chromosome	LOD Score	Trait	Location
5C_51000364	5C	7.524	Width	NERF
5C_51000364	5C	5.904	Width	Winner
5C_51399270	5C	5.950	Width	Winner
5A_137401828	5A	6.056	Width	SERF
5A_137401828	5A	6.272	Width	Volga
5A_137401828	5A	7.500	Width	Winner
5A_181475218	5A	6.325	Length	NERF
5A_181475218	5A	6.313	Width	Volga

Table 15. The significant markers and their closest RNA sequence, distance from the closest RNA sequence and if that sequence is found in other crops.

Marker	Closest RNA Sequence	Number of base pairs from the front of the RNA sequence	Gene sequence found in other crops?
5C_51000364	TRINITY_DN29674_c0_g1_i1	1,024 bp	no
5C_51399270	TRINITY_DN32700_c0_g1_i1	2,420 bp	Emmer
5A_137401828	TRINITY_DN111592_c0_g1_i1	56,284 bp	no
5A_181475218	TRINITY_DN101076_c0_g1_i1	98 bp	no

References

Andrie de Vries and Brian D. Ripley (2020). ggdendro: Create Dendrograms and Tree Diagrams Using 'ggplot2'. R package version 0.1.22. https://CRAN.R-project.org/package=ggdendro

Blast. (2021, March 17). Retrieved from National Center for Biotechnology Information: https://cran.r-project.org/web/packages/minque/index.html

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. (2007) TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633-2635.

Ching-Ting Huang, K. E.-F. (2020). Genome-Wide Association Study Reveals the Genetic Architecture of Seed Vigor in Oats. Genes | Genomes | Genetics, 4489-4503.

Douglas C. Doehlert, M. S. (2002). Sources of Variation in Oat Kernel Size. Cereal Chemistry, 528-534.

Douglas C. Doehlert, M. S.-L. (2004). Evaluation of Oat Kernel Size Uniformity. Crop Science, 1178-1186.

Endelman, J.B., and J.-L. Jannink. 2012. Shrinkage estimation of the realized relationship matrix. G3: Genes, Genomes, Genetics 2:1405-1413. doi: 10.1534/g3.112.004259

Flavio Breseghello, M. E. (2006). Association Mapping of Kernel Size and Milling Quality in Wheat Cultivars. Genetics Society of America, 1165-1177.

Francis H Webster, P. J. (2011). Oats Chemistry and Technology Second Addition. Minnesota: AACC International, Inc., 1-3

Giorgio Tumino, R. E.-J. (2017). Genome-wide association analysis for lodging tolerance

and plant height in a diverse European hexaploid oat collection. Euphytica, 163-165.

Hanna Haikka, O. M. (2020). Genome-wide Association Study of Genomic Prediction f for Fusarium graminearum Resistance Traits in Nordic Oat (Avena sativa L.). Agronomy, 174-197

J.S.S. Ammiraju, B. D. (2001). Identification of inter simple sequence repeat (ISSR) markers associated with seed size in wheat. Theoretical Applied Genetics, 726-732.

Felipe de Mendiburu and Muhammad Yaseen (2020). agricolae: Statistical Procedures for Agricultural Research. R package version 1.4.0 , https://myaseen208.github.io/agricolae/https://cran.r-project.org/package=agricolae.

Maryn O Carlson, G. M.-B.-L. (2019). Multivariate Genome-Wide Association Analysis Reveal the Genetic Basis of Seed Fatty Acid Composition in Oat (Avena sativa L.). Genes | Genomes | Genetics, 2963-2975.

Moran Nave, R. A.-Z. (2016). QTLs for uniform grain dimensions and germination selected during wheat domestication are co-located on chromosome 4B. Theoretical Applied Genetics, 1303-1315

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Song Yan, G. Z. (2011). Seed size is determined by the combinations of genes controlling different seed characteristics in rice. Theoretical Applied Genetics, 1173-1181.

Victoria Blake, M. W.-L. (2019). Grain Genes: centralized small grain resources and digital platform for geneticists and breeders. Database.

Jixiang Wu (2019). minque: Various Linear Mixed Model Analyses. R package version 2.0.0.https://CRAN.R-project.org/package=minque