Improvement of Milling and Nutritional Quality Characteristics of Oats Through Genomic Selection

Sudha Neupane Adhikari
South Dakota State University

Follow this and additional works at: https://openprairie.sdstate.edu/etd

Part of the Agriculture Commons, Genomics Commons, Molecular, Genetic, and Biochemical Nutrition Commons, and the Plant Sciences Commons

Recommended Citation
https://openprairie.sdstate.edu/etd/2146

This Thesis - Open Access is brought to you for free and open access by Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Open PRAIRIE: Open Public Research Access Institutional Repository and Information Exchange. For more information, please contact michael.biondo@sdstate.edu.
IMPROVEMENT OF MILLING AND NUTRITIONAL QUALITY
CHARACTERISTICS OF OATS THROUGH GENOMIC SELECTION

BY
SUDHA NEUPANE ADHIKARI

A thesis submitted in partial fulfillment of the requirements for the
Master of Science
Major in Plant Science
South Dakota State University
2017
IMPROVEMENT OF MILLING AND NUTRITIONAL QUALITY
CHARACTERISTICS OF OATS THROUGH GENOMIC SELECTION

SUDHA NEUPANE ADHIKARI

This dissertation is approved as a creditable and independent investigation by a candidate for the Master of Science degree in Plant Science 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-Treml, Ph.D.
Thesis Advisor

Date

David Wright, Ph.D.
Head, Department Plant Science

Date

Dean, Graduate School

Date
ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and humbleness to my major advisor Dr. Melanie Caffe-Treml, for providing me with this great opportunity to pursue my master's degree in the Department of Agronomy, Horticulture and Plant Science of such a renowned university. I am thankful for her constant supervision, valuable suggestions, scholastic guidance, continuous inspiration, and immense help during experiment, analysis and preparation of this manuscript.

I would like to express my sincere appreciation and respect to the member of my advisory committee and graduate school representative: Dr. Jixiang Wu, Dr. Jose Gonzalez, and Dr. Anna Sadovnikova for their cordial suggestions, constructive comments and valuable advice to complete my thesis. Dr. Jixiang Wu helped me throughout my master's degree. His guidance and encouragement also inspire me to learn more which helped me a lot in my research work and for the completion of my thesis. Also, Dr. Jose Gonzalez helped me a lot during the lab work and guiding me to use different lab equipment. His suggestions are also very helpful for my research work and thesis writing.

Special thanks go to Nick Hall who always helped me with my research activities. Also, I would like to thank Paul Richter from General Mills who has always helped me with my lab works. How can I forget my lab mate Bandana Dhungana, Jacob Baustian, Prakriti Sharma and graduate and undergraduate students Sarmila Katuwal Singh, Romancha K.C, Martee Larson who have helped me with the data collection and other various research activities. I express my regards and appreciations to all my friends here in South Dakota for their constant love and emotional support.
Lastly, I gratefully acknowledge my beloved parents, my supportive husband Amit Raj Adhikari, my brother, sisters and well-wishers for their countless blessings, spiritual support and sacrifice throughout my academic life.
TABLE OF CONTENTS

ABBREVIATIONS .............................................................................................................. vii
LIST OF TABLES ........................................................................................................ viii
LIST OF FIGURES ........................................................................................................ ix
ABSTRACT ...................................................................................................................... xi

CHAPTER 1 ...................................................................................................................... 1

1. Literature Review ...................................................................................................... 1
   1.1 Oats end-uses and production ........................................................................... 1
   1.2. Oat grain composition .................................................................................. 2
   1.3. Milling quality and its importance for the milling industry ....................... 2
      1.3.1. Groat Percent ......................................................................................... 3
      1.3.2. Thousand Kernel Weight ................................................................. 4
      1.3.3. Thin and Plump Kernels ...................................................................... 5
   1.4. Nutritional Quality ......................................................................................... 6
      1.4.1. β-Glucan ................................................................................................. 6
      1.4.2. Fat content ............................................................................................. 7
      1.4.3. Protein content ....................................................................................... 8
   1.5. Statistical modeling for taking account of field variation ......................... 8
   1.6. Genomic selection ......................................................................................... 10
      1.6.1. Principle of genomic selection .................................................................. 10
      1.6.2. Genomic selection vs. marker assisted selection (MAS) .................. 11
      1.6.3. Training population ............................................................................... 12
      1.6.4. Genotyping by sequencing (GBS) ...................................................... 12
      1.6.5. Genomic prediction methods .............................................................. 13

LITERATURE CITED .................................................................................................... 17

CHAPTER 2 ...................................................................................................................... 36

2. Genomic Selection for improving Key milling and Nutritional Quality traits in oats 36
   2.1. Introduction ................................................................................................... 36
   2.2. Material and methods .................................................................................. 39
      2.2.1. Breeding lines ......................................................................................... 39
      2.2.2. Phenotypes ............................................................................................. 39
      2.2.3. Genotyping .............................................................................................. 40
      2.2.4. Phenotypic data analysis ........................................................................ 42
2.2.5. Genomic Selection Analysis ................................................................. 43
2.2.6. Cross Validation ............................................................................ 44
2.3. Results and Discussion ................................................................ 44
  2.3.1. Comparison of statistical models for PYT data ......................... 46
  2.3.2. Prediction accuracy for milling and nutritional quality traits. ......... 47
  2.3.3. Comparison among genomic prediction methods ......................... 49
2.4. Conclusion ...................................................................................... 51
LITERATURE CITED ................................................................................ 70
ABBREVIATIONS

AVE= model averaging method
BLUP= best linear unbiased prediction
ELNET= elastic net
GBS= genotyping by sequencing
GEBV= genomic estimated breeding value
GS= genomic selection
MAS= marker assisted selection
minque= minimum norm quadratic unbiased estimation
NERF= Northeast Research Farm
PCR= polymerase chain reaction
PLSR= partial least square regression
PUFA= poly unsaturated fatty acid
PYT= preliminary yield trial
QTL= quantitative trait loci
REML= restricted maximum likelihood
RF= random forest
RRBLUP= ridge regression best linear unbiased prediction
SDSU= South Dakota State University
SERF= South East Research Farm
SFA= saturated fatty acid
SNP= single-nucleotide polymorphism
TP= training population
LIST OF TABLES

Table 2.1. Summary statistics for quality traits collected on grain samples from oat breeding lines evaluated in preliminary yield trials at four South Dakota locations in 2015 (n= 227) and 2016 (n=238)................................................................. 53

Table 2.2. Estimated variance components for oat milling and nutritional quality traits measured on 1080 oat samples collected from 227 genotypes grown at four South Dakota locations in 2015................................................................. 54

Table 2.3. Estimated variance components for oat milling and nutritional quality traits measured on 1080 oat samples collected from 238 genotypes grown at four South Dakota locations in 2016................................................................. 55

Table 2.4. Genotypic correlations among milling and nutritional quality traits in 2015 and 2016 (above and below diagonal respectively)................................................................. 56

Table 2.5. Average prediction accuracy for oat milling and nutritional quality traits using cross-validation in each environment for 227 oat breeding lines evaluated in seven environments................................................................. 56
LIST OF FIGURES

Fig. 1. 1. Oats production, consumption and import in US during 2013-2017 ............... 2

Fig. 2. 1. Flowchart of the process of Genotyping by Sequencing.............................. 41

Fig. 2. 2. Relationship between heritability and prediction accuracy for oat quality traits measured in the 2015 preliminary yield trial at four South Dakota locations. ............ 57

Fig. 2. 3. Relationship between heritability and prediction accuracy for oat quality traits measured in the 2016 preliminary yield trial at three South Dakota locations. ............ 57

Fig. 2. 4. Prediction accuracy for percent content using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.............................................................. 58

Fig. 2. 5. Prediction accuracy for β-glucan concentration using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.............................................................. 58

Fig. 2. 8. Prediction accuracy for percent thin kernels using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.............................................................. 60

Fig. 2. 9. Prediction accuracy for percent plump groat using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.............................................................. 60

Fig. 2. 10. Prediction accuracy for groat percent using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.............................................................. 61

Fig. 2. 11. Prediction accuracy for thousand kernels weight using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.............................................................. 61

Fig. 2. 12. Prediction accuracy for β-glucan using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations. .............................................................. 62

Fig. 2. 13. Prediction accuracy for fat content using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.............................................................. 62

Fig. 2. 14. Prediction accuracy for percent plump kernels using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.............................................................. 63
Fig. 2. 15. Prediction accuracy for percent thin kernels using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations ........................................................................................................ 63

Fig. 2. 16. Prediction accuracy for percent plump groat using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations ........................................................................................................ 64

Fig. 2. 17. Prediction accuracy for groat percent using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations ........................................................................................................ 64

Fig. 2. 18. Prediction accuracy for thousand kernels weight using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations ........................................................................................................ 65

Fig. 2. 19. Prediction accuracy for β-glucan and fat content using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations ........................................................................................................ 66

Fig. 2. 20. Prediction accuracy for groat percent and percent plump groat using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations ........................................................................................................ 67

Fig. 2. 21. Prediction accuracy for percent plump kernels and percent thin kernels using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations ........................................................................................................ 68

Fig. 2. 22. Prediction accuracy for thousand kernels weight using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations ........................................................................................................ 69
ABSTRACT

IMPROVEMENT OF MILLING AND NUTRITIONAL QUALITY
CHARACTERISTICS OF OATS THROUGH GENOMIC SELECTION
SUDHA NEUPANE ADHIKARI
2017

Oats can lower cholesterol, reduce risks of type-2 diabetes, and help prevent heart
diseases when consumed daily. Therefore, it is important to evaluate and select breeding
lines with desirable milling and nutritional quality traits. Genomic selection, which uses
genotyping data to predict the breeding value of an individual, is a promising method to
increase genetic gain by selecting for quality traits earlier in the line development
process. In this study, we collected phenotypic data for three nutritional traits (protein, β-
glucan, and fat content) and five milling quality traits (percent plumps, percent thins,
percent plump groat, groat percent and thousand kernel weight) on grain samples from
465 different oat genotypes grown at four locations in South Dakota in either 2015 or
2016. To take account of field variation, we investigated four linear mixed models with
the R package, minque, subject to the inclusion or exclusion of row and/or column
effects. Overall, inclusion of row and column effects reduced the error variance, and
accordingly, increased heritability and improve relative efficiency for most of the traits.
Thus, the full model with inclusion of row and column effects was applied to predict
genotypic effects for genomic selection analysis. All breeding lines were genotyped with
genotyping by sequencing (GBS) technique. Genomic selection models were evaluated
using five currently available genomic selection methods (RRBLUP, GAUSS, PLSR,
Elastic Net, and Random Forest) along with model averaging (AVE). Prediction accuracy
ranged from 0.19 to 0.74 and 0.30 to 0.70 among traits and locations in the year 2015 and 2016, respectively. Fat content and percent plump were the two traits with the highest prediction accuracy. β-glucan content, on the other hand, had the lowest prediction accuracy in both years. Overall the prediction accuracy was moderate to high for most of the traits in this study. Our results suggested that genomic selection could offer a valuable strategy to improve genetic gain for major milling and nutritional quality traits in oats.
CHAPTER 1

1. Literature Review

1.1 Oats end-uses and production

Oat (Avena sativa L.) is a multipurpose cereal crop cultivated over more than 9.5 million hectares globally (FAO, 2014). With an annual productivity of 2.3 ton ha\(^{-1}\) and a production of 22 million metric tons (FAO, 2014), it ranked sixth in terms of world cereal production after wheat, maize, rice, barley and sorghum (Hoffman, 1995; Kapoor and Batra, 2016, Serna-Saldivar, 2016). Oats are mostly grown between 35 to 60º latitude which corresponds to cool and moist climate. Due to the advancement in breeding program and the availability of high yielding and stress resistant cultivars, it is also gaining popularity in the southern subtropical regions (Sánchez-Martín et al., 2014). Oats are used primarily as livestock feed which accounted for 74% of total world's oat usages between 1995 and 2005 (USDA, 2015). However, in some countries like the US, its use as a human food is steadily increasing (Figure 1.1). US production is far less than total consumption and the majority of the oat consumed is imported from Canada (Figure 1.1). Oats contain high amount of dietary fibers, and phytochemicals and exhibit high nutritional value (Butt et al., 2008; Rasane et al., 2015). They have beneficial effects against gastrointestinal problems (Stark and Madar, 1994) as well as for the treatment of diabetes and cardiovascular disorders (Anderson 2003; Butt et al., 2008).
Fig. 1. Oats production, consumption and import in US during 2013-2017
(Source: COARSE GRAINS: WORLD MARKETS AND TRADE, USDA)

1.2. Oat grain composition

Oat grains are rich in beneficial nutrients like protein (11-15%), starch (59-70%), unsaturated fatty acids, soluble dietary fiber (12-15%), fat-soluble vitamins and β-glucans (3-7%) (Usman et al., 2010; Rasane et al., 2015; Sterna et al., 2016). The outer layer of tissues of the oat grain called bran is rich in minerals, vitamins, and antioxidant (gray et al., 2000). While soluble dietary fibers, minerals, and vitamins are concentrated on bran layers (Welch, 2011), proteins are stored in the endosperm and aleurone layer (Peterson, 2011).

1.3. Milling quality and its importance for the milling industry

Milling is the process of removing the husk from the grain (Winfield et al., 2007). Oats are regarded as whole-grain cereals as germ and bran fractions are not separated
from the endosperm during the milling process mainly due to the soft nature of groat (Welch, 2011). For the same reason and because of the distribution pattern of lipids in the seed, the milling process of oats is difficult in comparison to other cereals (Butt et al., 2008). The milling quality of oats is affected by their genetic make-up, the growing environment and how the grain was handled and stored. Groat percentage and size of the kernel affects the milling efficiency which are in turn associated to the genetic make-up of individual genotypes and are also influenced by the environment (Decker et al., 2014).

Generally, oat millers desire large kernel size because they produce larger flakes. Millers separate large and small kernels by setting standard grading procedure and thus they want uniformity in kernel size distribution across genotypes (Valentine et al., 2011). Therefore, selection of those genotypes producing uniformly large kernels in the early stage of breeding cycle is important for the milling industry.

1.3.1. Groat Percent

Groat percentage, ratio of groat (also called caryopsis) to the whole kernel, is one of the important milling quality traits of oats (Girardet and Webster, 2011). Groat is also referred as the husk less kernel and consists of germ, bran and endosperm. In general, groat contents ranges from 68-72% of the kernel (Webster, 1996) however there is potentiality of increasing the groat content up to 78-80% through breeding and appropriate growing conditions (Forsberg and Reeves, 1992). Generally, higher groat percentage is associated with the lager kernel size (Doehlert et al., 2006). However, this is not always true as some studies have shown that groat percentage obtained from heavier kernels to be on the lower side (Peek and Poehlman, 1949). Thus, in order to achieve the target of higher groat percentage, priority has been given towards increasing
the groat weight and decreasing hull weight. For this, different selections and breeding activities have been carried out (Stuthman and Granger 1977). Larger groats result in higher groat percentage (Doehlert et al., 2006). Peterson and Wood (1997) found a negative correlation between fat concentration and groat percent. Groat tends to be long and low in weight as concentration of fat increased. Whereas, β-Glucan was found to be positively correlated with the groat percent (Peterson et al., 1995). In some genotypes, β-Glucan is concentrated on outer portion of grain which can influence the grain volume to surface ratio (Miller and Fulcher 1994).

1.3.2. Thousand Kernel Weight

Weight of thousand undehulled grain at (12% moisture level) is another important parameter that provide some insight on physical properties of oats grain. We can use this parameter to estimate the milling quality of oats. Weight of oat grains influence test weight (weight per unit volume of grain at a standard moisture level) through the packing attribute of grain. Longer grains are found to reduce test weight, whereas plumper grain increase test weight (Doehlert, 2002). Thousand kernel weight and test weight both are used to predict the milling yield. However, test weight is found to be less efficient when we are trying to make milling yield prediction between genotypes (Girardet and Webster, 2011) mainly due to poor correlation between kernel size with the test weight. In a study involving ten oat genotypes, Doehlert (1999) found a correlation of r=0.542 between test weight and milling yield and this was low in comparison to milling yield prediction of 0.98 when only single genotype was considered (Peltonen-Sainio et al., 2004). Poor prediction of milling yields due to test weight when multiple genotypes were used might be due to variation in volume of empty spaces within the grains across the genotypes.
Which means, genotypes with tight fitting hull will have higher test weight in comparison with the genotype having loosely fit hull (Doehlert, 2006; Girardet and Webster, 2011). One of the reason for giving emphasis on thousand kernel weights over test weight is due to poor relation of kernel size with test weight (Stevens, 2006) which lead to poor prediction of milling yield. Several authors (Hutchinson, 1953; Johnson and Hartsing, 1963; Watson and heyne, 1977) emphasized thousand kernel weights to be more reliable estimator of milling yield in comparison to test weight.

### 1.3.3. Thin and Plump Kernels

Milling yield or milling quality of oats usually refers to the oats that are used as human food (Murphy and Winkler, 2017). Oats milling is done with the aim of enhancing the taste and quality of oats grain. Size of oat grains, more specifically, kernel length and thickness are usually taken as the measure for grading oats (Rasane et al., 2015) and have direct influence on the milling yield of oats and thus size is of great interest to milling industries (Doehlert et al., 2006). Kernel thickness refers to the two different grades of oats; thin kernel and plump kernel. Percent plump kernels and percent thin kernels are the percentage by weight of kernels remaining on 5.5/64 by ¾ inch and passing through a 5/64 by ¾ inch slotted screen after 30 strokes, respectively (Murphy and Winkler, 2017). The higher the fraction of thins, the poorer is the milling quality of oats because the thin kernels are too small to be processed and thus sorted out as byproducts for animal feed. (Girardet and Webster 2011). Therefore, to improve the milling quality of oats, the proportion of thins should be low. Likewise, Plump and Thin groat refer to the hull-less kernels (groat) that remain on the top of 5/64 by ¾ inch sieve and that passes through the same sieve after 30 strokes, respectively. There is a strong correlation between groat
weight and weight of grain (Doehlert and McMullen, 2000) and larger grains typically result in higher proportion of plump groat. Furthermore, for larger grains, smaller rotor speed is required to carry out dehulling in comparison to thin grains (Ganmann and Vorwerck, 1995) which means the higher the proportion of thin grains, the higher is the requirement for mechanical energy and the higher are the risks of grain breakage.

Usually, size distribution of oat grains is bi-modal, meaning there is presence of two different size grains in oat spikelet (Doehlert et al., 2004) where primary grains are larger and secondary grains are smaller (Doehlert et al., 2002). This proportion of thin grains to large grains in oat spikelet depends upon the genetic as well as environmental factors. Thus, it is desirable to select genotypes with larger grain size which should result in larger plump groats.

1.4. Nutritional Quality

1.4.1. β-Glucan

β-glucan, also called as lichenin is an important constituent of oats and is present in range of 3-7% (Usman et al., 2010). It is also called oat gum as they form a highly viscous solution when mixed with water (Welch, 1995). β-glucan is normally found in cell walls in the sub-aleurone layer (Wood et al., 1983; Fincher, 2009) and it is the size and distribution of cells that causes the variation in β-Glucan concentration among genotypes (Fucher and Muller 1993). β-Glucan found in oats is mainly linear polysaccharide (1→3), (1→4)-β-D-glucan, a soluble fiber that has been found effective for reducing cholesterol level in blood as well as maintaining the body weight (Wolever et al., 2010; Daou and Zhang, 2012). Furthermore, it is found effective in reducing the blood pressure level (Keenen et al., 2002), type-2 diabetes, and help prevent heart
diseases when consumed daily (Valentine and Cowan, 2004). Due to this health
benefiting characteristics, β-glucan has always been a selection criterion in oat breeding
programs (Newell et al., 2012). β-glucan concentration varies widely across oats
genotypes mainly due to environment and genetic make-up (Miller et al., 1993; Welch,
1995; Herrera et al., 2016) with genetic factor influencing the most (Lim et al., 1992;
Miller et al., 1993; Peterson et al., 1995). In a study conducted by Humphreys and Mather
(1996) heritability of β-glucan concentration ranged from 0.27 to 0.45.

1.4.2. Fat content

Fat, also known as oil or lipid is important nutritional quality of oat grains. Fat
gives twice the energy provided by carbohydrate and protein (Welch, 1995). Oat contains
substantial amount of polyunsaturated fatty acid (PUFA) along with fat soluble vitamin E
(Sterna et al., 2014). Oat grain contains the highest amount of fat (4.2-11.8 g/ 100 g) in
comparison to other cereals like rice (2.0-3.1 g/ 100 g), wheat (2.1-3.8 g/ 100 g) and
barley (3.3-4.6 g/ 100 g) (Zhou, 1999). There are numerous studies suggesting the
importance of PUFA on human health (Chillard et al., 2000; Gebauera et al., 2005) and
oats contain substantial amount (about 75%) of PUFA (Saastamoinen, 1989). According
to the WHO (2003), the ratio of PUFA to the saturated fatty acid (SFA) should be greater
than 0.4 for the food to be considered beneficial for human health. Based on this
guideline, oat is regarded as high quality nutritional food and thus are gaining popularity
among nutritionist. Fat content in oat grain is influenced by both environment and the
genetic make-up (Zhou, 1999). Fat content is a polygenic trait and it is found that the fat
concentrations in oats is highly heritable (Baker and McKenzie, 1972; Frey et al., 1975).
1.4.3. Protein content

Higher protein is a desirable characteristic of oats used for both food and feed purposes. Among the cereals, oat is valued as a low-cost protein source and has higher protein concentration than other cereals which provides considerable nutritive advantage (Peterson, 1992; Capouchova et al., 2004). Proteins in oats differ in their structural and distributional properties from other cereals (Rasane et al., 2015). In most cereals, the primary storage protein are prolamins (Klose and Arendt, 2012) which are limited in essential amino acid specifically lysine (Shewry, 2007). Whereas, primary storage protein in oats is globulin which contains higher concentration of lysine and other essential amino acids thus making it a superior protein source (Valentine and Cowan, 2004). Furthermore, composition of protein in oats is quite stable over a change in protein concentration in comparison to other cereals (Peterson, 1992). Protein content of the oats lies in the range between 9% and 24.5% (Robert et al., 1985; Mirmoghtadaie et al., 2009) and is influenced by the interaction between genotype and environment as well as N fertilizer (Welch and Young, 1980; Fan et al., 2009; Güler, 2011). To improve the protein concentration, emphasis has been given to breeding and selections.

1.5. Statistical modeling for taking account of field variation

Within field variation exists because of different factors like moisture gradient in the soil and variation in soil physical and chemical properties such as pH and fertility (Scharf & Alley, 1993; Adhikari et al., 1999; Wu & Dutilleul, 1999; Stroup, 2002). As a result, residual variation may be increased (Stroup, 2002; Wu et al., 2013) leading to the biased estimation of genetic effects (Bondalapati et al., 2015). Moreover, as block size increases, the precision in estimating residual variation could decrease (Aragaw, 2011).
Thus, emphasis should be given to the development of appropriate methods/models which could reduce the error variance for large field experiments.

Thus, the selection of an appropriate statistical model plays an important role in the proper analysis of field experiments and can minimize the lack of precision in estimated results (El-Mohsen, 2013). According to Santos et al. (2002), linear mixed model approach is one of the novel approaches for analyzing augmented experimental designs in which both row and column effects can be determined, and residual variance can be reduced accordingly (Wu et al., 2013). Linear mixed model approaches can be used for analysis of unbalanced data. Three different methods can be used to estimate variance components and predict random effects: maximum likelihood, restricted maximum likelihood, and minimum norm quadratic unbiased estimation (minque) (Hartley and Rao, 1967; Patterson and Thompson, 1971; Rao, 1971; Searle et al., 2009). Although the REML approach have been widely used and is the most popular one, in an experiment, minque approach was found comparable with the REML approach in terms of bias, testing power and Type I error (Nan et al., 2016). Moreover, the computational time for the minque package was low in comparison to REML approach (Nan et al., 2016). In addition, minque does not require the data to be normally distributed and iteration is not required for precise results while it is the case with REML approach (Rao, 1971). Minque Package (Wu, 2014) can also be integrated with a jackknife technique (a resampling technique) in order to test the significance of parameters (Wu et al., 2008; Wu et al., 2013) and this integration have been found useful for reducing standard error of estimated variance components thus raising the statistical power (Nan et al., 2016).
1.6. Genomic selection

From the very early days, people have been performing selection of different plant species mainly based on the plant’s appearance or phenotype. Today, to fulfill the rapidly increasing food demand of a rising population, we must look for breeding technologies that are more efficient in increasing genetic gain. New varieties must exhibit multiple characteristics such as high yield, disease and drought resistance, and good nutritional quality. Phenotypic selection can be quite challenging for low heritable polygenic traits which are usually influenced by the environment and G × E interactions (Bhat et al., 2016). The use of genotyping data to predict complex traits is showing great promise to improve the efficiency of the selection process.

1.6.1. Principle of genomic selection

Genomic selection consists in developing a prediction model which takes into account genetic effects from all markers distributed throughout the entire genome. Each marker effect is then added together to predict the breeding value of the individual (Boichard et al., 2016). Use of all markers allows GS to capture all the genetic variation for the trait (Goddarad and Hayes, 2009). Genomic prediction models can then be used to estimate the breeding value of the individuals that were not evaluated in the field and for which only genotypic data is available. Genomic prediction models are developed using a training population for which both genotypic and phenotypic data are available (Meuwissen et al., 2001). At first, marker effects in training set are estimated based on the response of marker allelic variance to the phenotypic performance and thus obtained marker effects are applied to the genotyped individuals of validation set to obtain GEBV. To obtain GEBV of individuals we can use equation:
\[ \text{GEBV}_j = \sum_{i=1}^{n} X_{ji} g_i \]  
(Solberg et al., 2008).

where \( \text{GEBV}_j \) = the GEBV of individual j; \( X_{ji} \) = the marker genotype of individual j; 
\( n \) = the number of markers; \( g \) = the estimates of the marker effects.

### 1.6.2. Genomic selection vs. marker assisted selection (MAS)

Marker Assisted Selection (MAS) in which molecular markers are used to select plants with desirable traits of interest (Bhat et al., 2016) has been very useful for plant breeders. However, MAS only consider a few significant markers (Lande and Thompson 1990) and focus is to identify major QTL (Asoro et al., 2011). Due to its inability to capture the smaller effect quantitative trait loci, MAS do not provide complete information for complex traits (polygenic traits) and sometimes is found to be inferior than conventional phenotype based selection method (Zhao et al., 2014). Thus, MAS has been confined to simpler traits (Bernardo, 2008; Xu and Crouch, 2008; Heffner et al., 2011).

Unlike MAS in which subset of markers with significant effects are identified, GS uses information from all the markers (Bernardo and Yu, 2007; Heffner et al., 2009; Asoro et al., 2011; Endelman, 2011). This avoid bias in estimation of marker effects. Heffner et al., (2010), while comparing between MAS and GS concluded that GS works better even for traits with low and moderate GEBV accuracies than MAS and for the traits with higher GEBV accuracies, the genetic gain is increased by several folds compared to MAS. Simulation study of Bernardo and Yu (2007) for the polygenic traits with low heritability in maize (Zea mays L.) showed that genome wide selection produced up to 43\% greater genetic gain than marker-assisted recurrent selection.
1.6.3. Training population

Training population (TP) comprises of individuals that have been genotyped and phenotyped and is used to develop and validate the prediction models (Meuwissen et al., 2011). Two different sets are obtained from TP; training sets and validation sets which are used to develop prediction models and obtain genomic estimated breeding value (GEBV). Since the training population requires phenotyping, one important consideration is to optimize TP size as we want to keep the phenotyping cost as minimum as possible (Cericola et al., 2017). Genomic selection accuracy is influenced by the number of lines in TP and markers used (Benardo and Yu, 2007). Although some researchers have found an increase in genomic prediction accuracy with increase in TP size (Zhong et al., 2009; Asoro et al., 2011; Lorenz et al., 2012), other studies suggested that the optimization and the use of small TP provide similar genomic prediction accuracies as large TP because the genomic prediction tends to reach a plateau although TP was increased (Isidro et al., 2015). It's also important to optimize the TP size based on the population for which we are predicting the GEBV. If the individuals in the prediction population are closely related to the individual in TP, even with the small TP and small number of markers, we can obtain an accurate prediction. While, if those populations are unrelated, we need to increase the size of TP and the number of markers used (Hickey et al., 2014).

1.6.4. Genotyping by sequencing (GBS)

Oat is polyploid in nature and thus genotyping is complicated due to the presence of homoeologous sub-genomes (Huang et al., 2014). GBS is a simple system which is highly multiplexed and is used to construct libraries for the next-generation sequencing
(Elshire et al., 2011). In polyploid crops like wheat, 10-20% increase in prediction accuracies was reported when GBS markers are used in comparison to the established marker platform (Poland et al., 2012). For finding SNPs and genotyping, GBS uses restriction enzymes which help to reduce genome complexity (Johnson et al., 2015). GBS is one of the most powerful, rapid and cost-effective technique to genotype breeding populations. In addition to this, GBS is effective for large scale study of genomic diversity, molecular marker and genomic selection (Zhao et al., 2014). It is gaining popularity in the field of plant breeding mainly due to characteristics like low cost, reduced sample handling, need for fewer PCR and purification steps, efficient barcoding and ease with which it can be scaled up (Davey et al., 2011; Elshire et al., 2011). Despite of these advantages, GBS needs great bioinformatics work because their adaptors are not specific for the end of the DNA fragment they bind. This leads to incomplete data (Fu, 2014) mainly due to low coverage sequencing (Davey et al., 2011) and sometimes duplication read (Anderson et al., 2017).

1.6.5. Genomic prediction methods

Various approaches have been proposed for GEBV prediction (Meuwissen et al., 2001; Gianola et al., 2006) and among them two approaches: BLUP and Bayesian are the most frequently used one (Su et al., 2010; Wang et al., 2012). BLUP and Bayesian methods differ in term of distribution for the variance of marker effects. BLUP assumes that effects of SNP are normally distributed with equal variance (Meuwissen et al. 2001) whereas in Bayesian method, each marker are allowed to have their own variance for allele effects which leads to unequal variance (Su et al., 2010; Asoro et al., 2011). Moser
et al. (2009), found accuracy for both BLUP and Bayesian to be quite similar with BLUP holding slight advantage over Bayesian due to minimal computational requirement.

Five different methods using BLUP have been commonly used to develop genomic prediction models and make estimation of the GEBV; Ridge regression best linear unbiased prediction (RRBLUP), GAUSS also known as reproducing kernel Hilbert space, Partial Least Square Regression (PLSR), Elastic Net (ELNET) and Random Forest (RF).

RRBLUP is one of the earlier methods proposed for the genomic selection (Whittaker et al., 2000; Meuwissen et al., 2001). Unlike general regression in which number of markers are limited to number of observations, ridge regression (RR) do not have such limitation and numerical stability is also high if the markers are highly correlated (Hoerl and Kennard, 2000). Higher numerical stability in RR is achieved because the coefficients of correlated predictor variables are shrink toward each other, allowing them to borrow strength from each other (Friedman et al., 2010). Because of these characteristics RR hold advantage over ordinary regression.

GAUSS is quite similar to RRBLUP, the only difference is that RRBLUP uses marker matrix to identify genetic covariance whereas GAUSS uses kernel effect based on Euclidean distance between genotypes to determine the genetic covariance (Endelman, 2011). GAUSS was found to have higher predictive ability than RRBLUP, RF, PLSR and ELNET when using cross validation (Battlefield et al., 2016). Likewise, Cuevas et al. (2017) found Gaussian kernel to have higher prediction accuracy than RRBLUP. Higher prediction ability of the Gaussian kernel models might be due to more flexible kernels.
that can effectively handle small and main effects of complex markers and marker-specific interaction effects (Cuevas et al., 2016).

Partial least square regression is similar to principal components regression (Mevik and Wehrens, 2007, Solberg et al., 2009). In comparison to PLSR, Bayesian method were found to be more accurate, however, PLSR was computationally simpler and faster than Bayesian technique (Solberg et al., 2009, Coster et al., 2010). Furthermore, PLSR is found to be effective method for genomic selection when multicollinearity exists among variables and when predictor matrix has more variables than observations (Colombani et al., 2012).

Like RRBLUP, a penalize method which shrink regression coefficient towards zero, Lasso is another penalized regression estimator. However, both of these techniques have some disadvantage. Ridge regression works well if there are many correlated predictor variables and all of them have non-zero coefficient whereas predictive ability of Lasso is not good when predictor variable is highly correlated (Waldmann et al., 2013). Thus, to overcome these demerits, ELNET was proposed by Zou and Hastie (2005) based on lasso and ridge regression penalties (Waldmann et al., 2013) that fits linear model with penalized maximum likelihood (Friedman, et al., 2009). Bühlmann and van de Geer (2011) found ELNET to have lower mean square error compared to RR and Lasso when predictor variables are highly correlated.

Likewise, RF is a machine learning non-parametric regression model having the higher potential to capture the non-additive effects (Heslot et al., 2012). It is based on a decision tree method (Breiman, 2001) where the prediction was made using 1000 trees.
Whenever we are unsure of the ideal prediction method or unknown about them, model averaging can be a valid option to quantify prediction accuracy (Raftery, et al., 1997, Raftery, et al., 2010). When we perform normal averaging, there is higher chance of overweighting average towards single model. Thus, to remove such biasness model averaging is performed by obtaining standardized values for mean and standard deviation (Battenfield et al., 2016). Since model averaging combines different models, it has potential to capture more genetic effects (Scutari, 2012). Hu et al. (2015) and Battenfield et al. (2016) found model averaging to provide higher prediction accuracy compared to any of the single model.
LITERATURE CITED


Battenfield, S. D., Guzmán, C., Gaynor, R. C., Singh, R. P., Peña, R. J., Dreisigacker, S., Fritz, A.K., & Poland, J. A. (2016). Genomic selection for processing and end-use
quality traits in the CIMMYT spring bread wheat breeding program. *The plant genome*, 9(2).


doi:10.3835/plantgenome2011.08.0024


their association to other traits by QTL analysis in two hexaploid oat populations. *Theoretical and Applied Genetics, 103*(1), 9-18.


the differences? Government of Western Australia, Department of Agriculture
and Food

properties of oat β-glucan influence its ability to reduce serum LDL cholesterol in
humans: a randomized clinical trial. *The American journal of clinical

interaction of cereal cell wall components with Congo Red and Calcofluor.
Specific detection and histochemistry of (1→ 3),(1→ 4)-β-D-glucan. *Journal of
Cereal Science, 1*(2), 95-110.

properties of oat fractions enriched in (1→ 3),(1→ 4)-β-D-glucan. *Cereal chemistry,
66*, 97-103.

Wu J (2014) minque: An R package for linear mixed model analyses. 1.1 edn,

[http://cran.rproject.org/](http://cran.rproject.org/)

Wu, J., Bondalapati, K., Glover, K., Berzonsky, W., Jenkins, J. N., & McCarty, J. C.
(2013). Genetic analysis without replications: model evaluation and application in


CHAPTER 2

2. Genomic Selection for improving Key milling and Nutritional Quality traits in oats

2.1. Introduction

Oats (*Avena sativa* L.) is an important cereal crop mainly because of its high content in dietary fibers like cellulose, arabinoxylans and soluble fibers (Zhou, et al., 1999; Drzikova, et al., 2005), phytochemicals, and because of its high nutritional value (Rasane et al., 2015). In addition, oats also contain high level of protein and unsaturated fats. Because of this health benefiting characteristics, oats are used to produce several food products including breakfast cereals, cereals bars, cookies, biscuits, probiotic drinks, and infant food. For those food items to be of superior qualities, oats should possess high nutritive and milling quality in addition to excellent agronomic characteristics.

Oat is polyploid and most of the quality traits in oats (milling and nutritional) are polygenic and complex in nature. Expression of such traits is governed by many genes with small to moderate genetic effects (Bhat et al., 2016). Low heritability and environmental sensitiveness further add complexity when selecting for such traits (Bhat et al., 2015). Selection for quality traits is often performed at the later stages of the varietal development process in conventional breeding program because milling quality determination is labor intensive and costly. Likely, breeding lines with desirable agronomic traits at early stages may be advanced but they could be then discarded at the later stages because of low quality. This process can be inefficient and result in extra costs which could be avoided if selection for quality traits was taking place at an earlier stage of the varietal development process.
In genomic selection (GS), the breeding value of an individual can be estimated based on genotyping data only. This technique could be considered as an effective approach to predicting breeding values for quality traits at an early stage of the breeding cycle. Although, GS is a relatively new approach in plant breeding program, numerous work has been carried out showing its potential as a successful technique for genetic improvement of crops. Crossa et al. (2010) compared GS and pedigree based models in wheat and maize and found that GS has higher predictive ability (7.7 to 35.7% higher) than pedigree based selection. Similarly, Massman et al. (2013) also carried out GS in maize and reached to the conclusion that GS leads to higher genetic gain compared to MAS. Battenfield et al., (2016) regarded GS as a powerful technique to enhance early generation selection efficiency for quality traits including test weight, thousand kernel weights, protein content and flour yield in wheat. Fiedler et al. (2017) stated that GS was an effective tool for genetic improvement of durum wheat with prediction accuracies ranging from 0.27 to 0.66. Even though GS doesn’t replace completely the need for field evaluations (Storlie et al., 2013), it can help to predict how well a test line performs before it is planted in the field and thus may decrease high phenotyping costs (Hayes et al., 2009).

Genomic selection accuracy is influenced by the number of lines in TP and markers used (Benard and Yu, 2007). Although some researchers have found an increase in genomic prediction accuracy with increase in TP size (Zhong et al., 2009; Asoro et al., 2011; Lorenz et al., 2012), other studies suggested that the optimization and the use of small TP provide similar genomic prediction accuracies as large TP because
the genomic prediction tends to reach a plateau as the size of the TP increased (Isidro et al., 2015).

Because the size of the training population is often large, when collecting phenotypic data on the training population, it may not be feasible economically for the researcher to follow a replicated experimental design. Furthermore, limited space and seeding materials are some key constraints which propels researcher towards non-replicated augmented experimental designs. In such designs, only standard or check lines are replicated for controlling experimental error (Santos et al. 2002). Under such scenario, addressing within field variation can be quite challenging. Therefore, when analyzing field trials with such designs, the probability of drawing biased conclusion increases which could impact selection efficiency. Thus, we need to develop proper statistical model (post data treatments) with the inclusion of moving mean or row-column effects (Peiris et al., 2008, Mullar et al., 2010, Leiser et al., 2012) for the analysis of phenotypic data (Lado et al., 2015). This will improve the prediction accuracy by addressing field variation (if present) for obtaining more reliable phenotypic value.

The objectives of this project are to: 1. Identify the best statistical model to analyze preliminary yield trials which follow an augmented design; and 2. Evaluate the potential of GS for the improvement of milling and nutritional quality traits. To address our objectives, we have considered eight different milling and nutritional quality traits of oats (percent protein, β-glucan, fat content percent plump kernels, percent thin kernels, percent pump groat, and thousand kernels weight). Our main hypothesis was that genomic selection models can be developed to predict milling and nutritional quality traits with sufficient levels of accuracy to be used in oat breeding programs.
2.2. Material and methods

2.2.1. Breeding lines

The breeding lines from the 2015 (227 lines) and 2016 (238 lines) preliminary yield trials (PYT) of the South Dakota State University oat breeding program were used for this study. Among these lines 15 lines (including four checks) were common between two years of trials evaluation.

2.2.2. Phenotypes

In 2015, the 227 lines were grown at four different locations: Northeast Research Farm (NERF) located in Southshore, SD (45°6′18″N and 96°55′41″W), the Southeast Research Farm (SERF) located near Beresford, SD (43°4′51″N and 96°46′34″W), near Volga, SD (44°19′19″N and 96°55′28″W), and near Winner, SD (43°22′26″N and 99°51′28″W). The experiment at each location was conducted without replications except that four checks (Horsepower, Hayden, Natty and Shelby 427) were replicated for nine times. Each trial had rectangular arrangement with six columns and fifteen rows in NERF and SERF, five columns and eighteen rows in Volga, and eight columns and eleven rows in Winner. The plot size was 5 by 6 feet at all locations except for Winner where it was 5 by 13 feet. In 2016, the 238 breeding lines were planted at the same four locations, but no plot were harvested at SERF due to excessive rainfall following planting.

For all samples (n=1890), phenotypic data were collected for several traits of interest to the milling industry (percent plump kernels, percent thin kernels, percent plump groat, groat percent, thousand kernels weight) and three nutritional quality traits (percent protein, β-Glucan, and fat content). Percent plump and percent thin kernels are the percentage by weight of kernels remaining on 5.5/64 by ¾ inch and passing through a
5/64 by 3/4 inch slotted screen, respectively (E.L. Erickson products, Brookings, South Dakota) after 30 strokes. Percent plump groat is the ratio of plump groat weight (groat remaining on 5/64 by 3/4 inch slotted screen after 30 strokes) to the total groat weight expressed in percentage. Groat percent is the ratio of the groat weight to the grain weight measured after running the grains through a Codema -Laboratory Oat Huller (Codema, Minnesota). Thousand kernels weight is the weight of 1000 grains. β-glucan and fat content were determined by NIR spectroscopy on ground groat flour while percent protein on the groat. All data collection for year 2015 samples was done by General Mills (Le Sueur, MN), quality data collection for year 2016 samples was performed at SDSU.

2.2.3. Genotyping

Breeding lines (465) from the year 2015 and 2016 were genotyped. For that leaf tissue was collected from a single plant and DNA was extracted at the USDA-ARS genotyping lab in Fargo, North Dakota. DNAs were quantified, normalized and then digested with a two-enzyme approach, then they were barcoded, pooled, cleaned, amplified and sequenced (Poland, et al., 2012). These breeding lines were genotyped at the USDA-ARS genotyping lab in Fargo, ND and SNP calling was done by Dr. Tinker’s lab using the UNEAK pipeline (Huang et al., 2014). Markers data were downloaded from T3/Oat database(https://triticeaetoolbox.org/oat/genotyping/genotype_selection.php). Filtering was done on T3/oat prior to downloading; markers with greater than 20% missing data and less than 5% minor allele frequency were discarded. This resulted in 40,180 markers for year 2015 and 38,329 markers for year 2016. Some of the breeding lines used in 2016 were already genotyped in 2015 thus we needed to select common markers for both 2015 and 2016 to include those common breeding lines in 2016. This
inclusion of common breeding lines resulted in 26,276 markers for 2016. We obtained these markers in allelic form which was then converted to numerical format (-1, 0, 1) using the GAPIT package in R (Lipka et al., 2012) where, -1, 0, and 1 represents homozygous recessive, heterozygous and homozygous dominant, respectively. All missing alleles were imputed as heterozygous.

Fig. 2. 1. Flowchart of the process of Genotyping by Sequencing. Adapted from “Genotyping-by-sequencing for plant breeding and genetics” by J. Poland and T.W. Rife, 2012, *The Plant Genome*, 5(3), 92-102
2.2.4. Phenotypic data analysis

Phenotypic data were analyzed using a linear mixed model approach with the R package minque (Wu, 2014). Four linear mixed models were compared for computing Best Linear Unbiased Predictor (BLUP) values for each genotype as follow:

Model 1: most reduced model
\[ y = \mu + L + G + e \]

Model 2: including row effect
\[ y = \mu + L + G + R(L) + e \]

Model 3: including column effect
\[ y = \mu + L + G + C(L) + e \]

Model 4: full model including row and column effects
\[ y = \mu + L + G + R(L) + C(L) + e \]

Where, y is an observation; \( \mu \) is the population mean; L is a location effect; G is a genotypic effect; R is a row effect nested in location; C is a column effect nested in location, and e is random error. Among these four models, model 1 is the most reduced model without row or column effect; model 2 includes row effect; model 3 includes column effect; and model 4 includes both row and column effects and is considered as a full model in this study. We randomly divided 20 groups for the jackknife process. All components except the population mean in these four models were treated as random.

To identify the best model for computing BLUP for developing genomic selection models the four statistical models were compared based on broad sense heritability and relative efficiency:

\[
\text{Broad sense heritability} (H^2) = \frac{\text{Variance due to genotype} (V_G)}{\text{Variance due to genotype} (V_G) + \text{Variance due to error} (V_e)}
\]
Relative Efficiency (R.E.) = \frac{Error variance of most reduced model}{Error variance of model with row and (or) column effect}

The greater the heritability and relative efficiency, the more efficient the model is.

Data were analyzed for single locations and across locations in each year. BLUPs were computed using the model with the greatest efficiency.

### 2.2.5. Genomic Selection Analysis

BLUP (best linear unbiased prediction) values obtained from the best linear mixed model and genotypic data for breeding lines from each year were used to compute GEBV using different genomic selection methods; ridge regression best linear unbiased prediction (RRBLUP), GAUSS also known as reproducing kernel Hilbert space, partial least square regression (PLSR), elastic net (ELNET) and random forest (RF). Apart from these methods, we used average modeling method (AVE) that normalizes mean and standard deviation of GEBV obtained with the five other models to obtain prediction accuracy. Packages to conduct these methods were developed in R programming platform (R Development Core Team, 2014). RRBLUP and Gaussian kernel (GAUSS) methods were carried out using ‘rrBLUP’ package in R, as described in Endelman (2011). Remaining three GS methods; PLSR, ELNET and RF were operated using R packages ‘pls’ (Mevik and Wehrens, 2007), ‘glmnet’ (Friedman, et al., 2009), and ‘randomForest’ (Liaw and Wiener, 2002), respectively. These GS methods were combined in a single package called ‘GSwGBS’ by Gaynor (2015). More details regarding these methods are presented in literature review section (2.2.5).
2.2.6. Cross Validation

GS methods were evaluated through cross validation (CV) technique. The data (BLUP values obtained from the best linear mixed model and genotypic data) was divided into five equal subsets (fivelfold cross-validation) and 80% of the data (four subsets) was used as training set. The remaining 20% (one subset) was used to validate the model. The cross-validation was repeated 20 times. Prediction accuracy for each fold was obtained by plotting the correlation between genomic estimated breeding value (GEBV) and the BLUP value (R Development Core Team, 2014) for all GS methods.

2.3. Results and Discussion

Mean, variance and range of eight different milling and nutritional quality traits (percent plump kernels, percent thin kernels, percent pump groat, thousand kernels weight, percent protein, β-glucan, and fat content) for 227 breeding lines evaluated at four locations in 2015 and 238 breeding lines evaluated at three locations in 2016 are presented in Table 2.1. For the three nutritional quality traits of oats, mean and variance look quite similar across locations in both year 2015 and 2016. For all other milling traits except groat percent, the overall variance increased in year 2016 compared to year 2015.

Heritability for most of the traits in our study was moderate to high in both year 2015 and 2016 (Table 2.2 and Table 2.3). β-glucan had the lowest heritability among the traits evaluated (0.37 to 0.44 in 2015 and 0.33 to 0.36 in 2016) whereas fat content had high heritability (0.77 to 0.81). Heritability estimates obtained in our study for β-glucan were in accordance with a study conducted by Humphreys and Mather (1996) where heritability ranged from 0.27 to 0.45. Estimated heritability for groat percent in our study ranged from 0.33 to 0.36 in 2015 and 0.40 to 0.43 in 2016. Humphreys and Mather
(1996) found an estimated heritability for groat percent ranging from 0.23 to 0.32. However, Stuthman and Granger (1977) found wider heritability range for groat percent (0.34 to 0.72). Results from these studies including ours suggest that groat percent is largely influenced by the environment. For most of the traits, the heritability estimates were consistent in both years of the study except for percent thin kernels and percent plump groat. Heritability was 12% and 16% lower for percent plump groat and percent thin kernels respectively, in 2016 compared to 2015. This decrease in heritability was mainly due to an increase in error variance 74% for percent plump groat (increased from 61.8 to 242) and 58% for percent thin kernels (increased from 11.0 to 26.6) (Table 2.2 and 2.3).

Pearson correlation coefficients among the traits for both year 2015 and 2016 are provided in Table 2.4. As expected, percent plump kernels and percent plump groat were highly positively correlated (r = 0.81 in 2015 and r = 0.83 in 2016) and both were highly negatively correlated with percent thin kernels. Thousand kernels weight was positively correlated with percent plump kernels and percent plump groat and negatively correlated with percent thin kernels. Again, this was expected as plumper kernels are often heavier. Groat percent was significantly correlated to percent plump groat but the strength of the correlation was relatively low (r = 0.41 in 2015 and r = 0.27 in 2016). Groat percent is likely to be more dependent on the groat and hull weight than on the size of the groats. Groat percent was not significantly correlated with percent plump kernels or percent thin kernels. β-glucan was not correlated with any other traits in 2015 but it was positively correlated with fat content in 2016 (r = 0.30). Peterson et al. (2005) and Bleidere et al. (2014) found a similar positive association between β-glucan and fat content however
Mut et al. (2016) found that β-glucan was negatively associated with fat content. Peterson et al. (1995) concluded that the genotypic variation of β-glucan was sufficient for breeding progress. Percent protein was only evaluated in 2015 and was weakly but significantly correlated with β-glucan ($r = 0.14$), fat content ($r = 0.16$), percent plump kernels ($r = 0.20$), percent thin kernels ($r = -0.23$), groat percent ($r = 0.30$), percent plump groat ($r = 0.33$), and thousand kernels weight ($r = 0.28$). Similar positive association between crude protein and β-glucan was reported previously (Bleidere et al., 2014; Mut et al., 2016). Likewise, Magliano et al. (2014) found percent protein to be negatively associated with percent thin kernels as observed in our study.

### 2.3.1. Comparison of statistical models for PYT data

Estimated variance components obtained for four different models (Models 1-4)) are listed in Table 2.2 and Table 2.3. Row and column effects were significant for almost all the milling and nutritional quality traits except for β-glucan and percent plump kernels. Although, variance contributed by row and column effects was lower than variance due to location and genotypes in all models, there was a decrease in error variance when row and column effects were included. This finding was in agreement with Bondalapati et al. (2015). Among the four different models, residual variance was the highest for model 1 followed by model 2 and model 3, suggesting the presence of spatial variation (Table 2.2 and Table 2.3). Residual variance was lowest using model 4 for all traits in 2015 and 2016. Based on the estimated error variances from different models, model 4 (full model) was either more efficient or comparable to other models depending on the traits in 2015 and 2016.
Heritability increased as we included row and/or column effects in models and was highest in model 4 for most of these traits. Effectiveness of row and/or column effects inclusion was supported by the increased relative efficiency of the full model (model 4) in comparison to most reduced model (Table 2.2 and Table 2.3). To identify the best model, relative efficiency, error variance, and heritability were considered. The statistical model with the lowest error variance and the highest relative efficiency and heritability was model 4 (including both row and column effect) for all traits when locations were combined. Model 4 was therefore found to be superior in comparison to other models. However, when locations were analyzed individually, the best model varied depending on the traits and the location (data not shown).

2.3.2. Prediction accuracy for milling and nutritional quality traits.

Five different genomic selection (GS) methods; ELNET, RF, GAUSS, RRBLUP, PLSR and AVE (Model Averaging by using normalized mean and standard deviation of individual method) were used to estimate the prediction accuracy for eight different milling and nutritional quality traits of oats in each location for 2015 and 2016.

Prediction accuracy obtained using different GS methods for percent protein, β-glucan, fat content, percent plump kernels, percent thin kernels, percent plump groat, groat percent and thousand kernels weight for year 2015 is graphically presented in Fig. 2.4 to 2.11. Similarly, prediction accuracy for those traits (except protein percent) for 2016 are presented in Fig. 2.12 to 2.18. Higher prediction accuracy was obtained for fat content (0.46 to 0.73 in 2015 and 0.54 to 0.70 in 2016), percent plump kernels (0.65 to 0.74 in 2015 and 0.44 to 0.62 in 2016) and percent thin kernels (0.58 to 0.72 in 2015 and 0.52 to 0.61 in 2016). Whereas, lower prediction accuracy was obtained for β-glucan
(0.19 to 0.43 in 2015 and 0.30 to 0.45 in 2016) (Table 2.5). Overall prediction accuracies were moderate to high for most of the traits under study. Asoro et al. (2011) reported a similar range of prediction accuracy for β-glucan (0.35 to 0.47) and groat percent (0.21 to 0.30) using two different GS methods and various training population size. Although the literature on the use of GS for predicting quality traits of oats is scarce, several GS studies have reported the use of GS to predict end-use quality traits in wheat. Moderate prediction accuracy for wheat protein content (0.25 to 0.51) was obtained using RR-BLUP (Michel et al., 2016). Lado et al. (2017) reported a prediction accuracy of 0.63 for wheat protein content using Bayesian lasso. Fiedler et al. (2017) used GS methods for cross validation and forward prediction approach and obtained prediction accuracy ranging from 0.18 to 0.56 for wheat flour protein content. Prediction accuracy for wheat’s thousand kernels weight ranged from 0.75 to 0.84 (Lado et al., 2013). Similar moderate GS accuracy for protein content (0.3 to 0.5) and thousand kernels weight (0.22 to 0.5) in wheat was found by Battenfield et al. (2016) across various genomic selection methods.

Prediction accuracy of all the GS methods for percent protein, β-glucan, percent plump groat, percent thin kernels and thousand kernels weight decreased in 2016 compare to 2015 (Fig. 2.19 to Fig. 2.22). This decrease in predictive ability might be due to higher influence of environment and low estimated heritability of those traits (Table 2.3 and 2.4). In both year 2015 and 2016, highest prediction accuracy values were obtained for fat content and percent plump kernels followed by moderate prediction accuracy for percent protein, thousand kernels weight and percent plump groat. In 2015, prediction models for traits with higher heritability had higher accuracy ($R^2=0.58$). In 2016 however, the correlation between heritability and prediction accuracy was low.
(R²=0.21). For fat content, prediction ability did not differ much across the years (Fig. 2.19) and this might be due to high heritability of fat content (Table 2.2 and Table 2.3). However, for groat percent, prediction accuracy of all GS methods increased in 2016 compared to 2015 (Fig. 2.20). Prediction accuracy was lower (0.42 to 0.48) in 2015 where heritability ranged from 0.30 to 0.36. Whereas in 2016, prediction accuracy ranged from 0.41 to 0.63 and the heritability ranged from 0.40 to 0.43. Similar results showing an association between prediction accuracy and heritability was previously reported (Lorenzana and Bernardo., 2009; Zhong et al., 2009; Muranty et al. ,2015).

2.3.3. Comparison among genomic prediction methods

All genomic selection methods used in this study provided good prediction accuracy in both 2015 and 2016. Overall, GAUSS followed by AVE were the prediction methods with the highest prediction accuracy for the majority of the traits. PLSR had low prediction accuracy in comparison to other methods. Similar results suggesting higher prediction accuracy of Gaussian kernel over other methods were reported previously (Endelman, 2011; Battenfield et al., 2016; Cuevas et al., 2017). Whenever we are unsure of the ideal prediction method or unknown about them, model averaging can be a valid option to quantify prediction accuracy (Raftery, et al., 1997, Raftery, et al., 2010). For AVE, normalized mean and standard deviation for each entry of prediction models were used which resulted in consistently high prediction accuracy for most of the traits and across locations and year (Battenfield et al., 2016).

Prediction accuracy of different GS methods for percent protein did not varied significantly (Fig 2.4). Though not statistically different, AVE, ELNET and GAUSS resulted in higher prediction accuracy across environments.
For β-glucan, prediction accuracy did not vary significantly among GS methods in 2015 (Fig. 2.5). However, in 2016, there were statistically significant differences in predictive ability (Fig. 2.12). AVE, RF and GAUSS performed better in comparison to other methods.

For fat content, there was statistical difference among GS methods in NERF, SERF and Volga where predictive ability of GAUSS, PLSR and RRBLUP were significantly lower (Fig. 2.6). Similarly, for 2016, all other GS methods resulted in statistically similar prediction accuracy except PLSR (Fig 2.13). For percent plump kernels, GAUSS, RRBLUP and AVE resulted in higher prediction accuracy in both years (Fig. 2.7 and Fig. 2.14). Likewise, for percent thin kernels, predictive ability of GS methods varied significantly in 2016 with RF providing statistically lower prediction accuracy (Fig 2.15). Overall, GAUSS and AVE provided higher prediction accuracy in both years.

Prediction accuracies of different GS methods for percent plump groat differed statistically in 2015 (Fig 2.9). However, for 2016, we did not find any statistically significant differences between GS methods (Fig 2.16). In both years, GAUSS and AVE provided higher prediction accuracy whereas RF and PLSR prediction accuracy were lower.

There was a significant difference in prediction accuracy among GS methods for groat percent in 2015 (Fig. 2.10). GAUSS, AVE and RRBLUP resulted higher prediction accuracy. For thousand kernels weight, GS methods PLSR, AVE and RRBLUP yielded
in higher prediction accuracy and RF was the least performing methods for both years (Fig. 2.11 and Fig. 2.18).

2.4. Conclusion

Oat milling, and nutritional quality traits are important to producers, the food industry and consumers. However, such traits are governed by many genes thus making it difficult to carry out efficient selection using standard breeding method. It is not economically feasible to phenotype thousands of breeding lines for all quality traits in early stages of breeding. Therefore, selection for such traits are typically done at later stages and this may lead to the advancement of the lines with poor quality. GS could be very useful for improving the selection efficiency for those traits by selecting at early stage. Furthermore, GS could help shorten the breeding cycle as individuals with higher genomic estimated breeding value (GEBV) can be selected as parental lines.

Micro-environmental variations at the testing site might affect the precision of the measured phenotype, and it is essential to account for those non-genetic variation (field variation) for obtaining more precise phenotypic data and ultimately more accurate GS prediction models. To take account of field variation, we compared four statistical models in minque package of R that were based on inclusion of row and/or column effects. Inclusion of row and column effects reduced the error variance, increased heritability and improve relative efficiency for most of the traits. Thus, we used model 4 (Full model, with row and column effect) to compute BLUP for the training and validation sets for genomic selection.

Prediction accuracy ranged from 0.19 to 0.74 and 0.30 to 0.70 for the eight traits across locations in year 2015 and 2016, respectively. Overall the prediction accuracy was
moderate to high for most of the traits under study with higher prediction accuracy for fat content and percent plump kernels and low for $\beta$-glucan. We found positive association between heritability and prediction accuracy for different GS methods. All genomic selection methods used in this study tend to provide good prediction accuracy in both year 2015 and 2016. Overall, GAUSS was the best prediction method for most of the traits whereas PLSR had low prediction accuracy in comparison to the other methods. AVE resulted in consistently high prediction accuracy for most of the traits across locations and years. Overall, the levels of prediction accuracy obtained for the different milling and nutritional quality trait was sufficient to serve as a valuable technique for improving milling and nutritional quality traits of the oats.

However, genomic prediction models were developed for each year separately and models were trained and validated using the genotypic and phenotypic data from the same year. Such approach can lead to a biased estimation of prediction accuracy as genotype by year interaction is not estimated (Albrecht et al. 2014, Wang et al. 2014). It will be important to use additional years of data and to develop prediction models that encompass several years and evaluate how well prediction models developed based on previous years will be able to predict GEBV of following years.

The next step in this study should be to test validation scheme such as forward prediction where breeding lines from the previous year are used to predict GEBV for breeding lines in the following year. To predict 2016 PYT lines, we should use 2015 PYTs as training population and to predict 2017 lines, we should use 2015 and 2016 PYTs as training population. This should result in increased and more stable prediction accuracy (Muir, 2007).
Table 2.1. Summary statistics for quality traits collected on grain samples from oat breeding lines evaluated in preliminary yield trials at four South Dakota locations in 2015 (n= 227) and 2016 (n=238)

<table>
<thead>
<tr>
<th>Environment</th>
<th>Percent Protein</th>
<th>β-glucan</th>
<th>Fat content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Range</td>
</tr>
<tr>
<td>2015NERF</td>
<td>13.9</td>
<td>1.4</td>
<td>9.8</td>
</tr>
<tr>
<td>2016NERF</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2015SERF</td>
<td>12.9</td>
<td>0.9</td>
<td>10.5</td>
</tr>
<tr>
<td>2015Volga</td>
<td>11.6</td>
<td>0.6</td>
<td>10.1</td>
</tr>
<tr>
<td>2016Volga</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2015Winner</td>
<td>14.3</td>
<td>1.6</td>
<td>10.9</td>
</tr>
<tr>
<td>2016Winner</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environment</th>
<th>Percent Plump Kernels</th>
<th>Percent Thin Kernels</th>
<th>Percent Plump Grain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
<td>Range</td>
</tr>
<tr>
<td>2015NERF</td>
<td>71.2</td>
<td>217.9</td>
<td>29.4</td>
</tr>
<tr>
<td>2016NERF</td>
<td>25.4</td>
<td>288.5</td>
<td>3.2</td>
</tr>
<tr>
<td>2015SERF</td>
<td>67.9</td>
<td>245.5</td>
<td>24.6</td>
</tr>
<tr>
<td>2015Volga</td>
<td>60.2</td>
<td>260.3</td>
<td>18.8</td>
</tr>
<tr>
<td>2016Volga</td>
<td>33.8</td>
<td>372.4</td>
<td>3.2</td>
</tr>
<tr>
<td>2015Winner</td>
<td>79.0</td>
<td>100.0</td>
<td>41.7</td>
</tr>
<tr>
<td>2016Winner</td>
<td>48.0</td>
<td>333.2</td>
<td>5.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environment</th>
<th>Groat percent</th>
<th>Thousand Kernels Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Variance</td>
</tr>
<tr>
<td>2015NERF</td>
<td>74.4</td>
<td>11.9</td>
</tr>
<tr>
<td>2016NERF</td>
<td>67.3</td>
<td>9.4</td>
</tr>
<tr>
<td>2015SERF</td>
<td>70.7</td>
<td>15.3</td>
</tr>
<tr>
<td>2015Volga</td>
<td>68.2</td>
<td>33.5</td>
</tr>
<tr>
<td>2016Volga</td>
<td>67.6</td>
<td>23.5</td>
</tr>
<tr>
<td>2015Winner</td>
<td>69.4</td>
<td>10.2</td>
</tr>
<tr>
<td>2016Winner</td>
<td>73.6</td>
<td>9.6</td>
</tr>
</tbody>
</table>
Table 2.2. Estimated variance components for oat milling and nutritional quality traits measured on 1080 oat samples collected from 227 genotypes grown at four South Dakota locations in 2015.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Percent Protein</th>
<th>β-Glucan</th>
<th>Fat Content</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
</tr>
<tr>
<td>$V_c$</td>
<td>1.47***</td>
<td>1.46***</td>
<td>1.47***</td>
</tr>
<tr>
<td>$V_g$</td>
<td>0.45***</td>
<td>0.46***</td>
<td>0.44***</td>
</tr>
<tr>
<td>$V_{(L)}$</td>
<td>0.04***</td>
<td>0.04***</td>
<td>0.04***</td>
</tr>
<tr>
<td>$V_{(G)}$</td>
<td>-</td>
<td>0.16***</td>
<td>0.16***</td>
</tr>
<tr>
<td>$V_{(L)}$</td>
<td>0.67***</td>
<td>0.62***</td>
<td>0.54***</td>
</tr>
<tr>
<td>$H^2$</td>
<td>0.40</td>
<td>0.42</td>
<td>0.45</td>
</tr>
<tr>
<td>RE</td>
<td>-</td>
<td>1.07</td>
<td>1.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Percent Plump Kernels</th>
<th>Percent Thin Kernels</th>
<th>Percent Plump Groat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
</tr>
<tr>
<td>$V_c$</td>
<td>60.8***</td>
<td>60.1***</td>
<td>61.5***</td>
</tr>
<tr>
<td>$V_g$</td>
<td>138***</td>
<td>137***</td>
<td>134***</td>
</tr>
<tr>
<td>$V_{(L)}$</td>
<td>2.69***</td>
<td>-</td>
<td>2.72***</td>
</tr>
<tr>
<td>$V_{(G)}$</td>
<td>-</td>
<td>2.65***</td>
<td>2.82***</td>
</tr>
<tr>
<td>$V_{(L)}$</td>
<td>60.7***</td>
<td>59.1***</td>
<td>60.2***</td>
</tr>
<tr>
<td>$H^2$</td>
<td>0.70</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td>RE</td>
<td>-</td>
<td>1.03</td>
<td>1.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Groat Percent</th>
<th>Thousand Kernels Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
</tr>
<tr>
<td>$V_c$</td>
<td>7.09***</td>
<td>7.09***</td>
</tr>
<tr>
<td>$V_g$</td>
<td>5.03***</td>
<td>5.06***</td>
</tr>
<tr>
<td>$V_{(L)}$</td>
<td>0.42***</td>
<td>-</td>
</tr>
<tr>
<td>$V_{(G)}$</td>
<td>-</td>
<td>2.21***</td>
</tr>
<tr>
<td>$V_{(L)}$</td>
<td>11.9***</td>
<td>11.6***</td>
</tr>
<tr>
<td>$H^2$</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>RE</td>
<td>-</td>
<td>1.02</td>
</tr>
</tbody>
</table>

*Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability

M1: most reduced model (Model 1), M2: includes row effect (Model 2), M3: includes column effect (Model 3), and M4: includes both row and column effects (Model 4), $V_c$: variance due to location, $V_g$: variance due to genotype, $V_{(L)}$: variance due to row nested in location, $V_{(G)}$: variance due to column nested in location, $H^2$: broad sense heritability, RE: relative efficiency
Table 2.3. Estimated variance components for oat milling and nutritional quality traits measured on 1080 oat samples collected from 238 genotypes grown at four South Dakota locations in 2016.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>β-Glucan</th>
<th>Fat Content</th>
<th>Percent Plump Kernels</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
</tr>
<tr>
<td>(V_L)</td>
<td>0.03**</td>
<td>0.03**</td>
<td>0.03**</td>
</tr>
<tr>
<td>(V_G)</td>
<td>0.09***</td>
<td>0.10***</td>
<td>0.09***</td>
</tr>
<tr>
<td>(V_{RI})</td>
<td>7.05</td>
<td>-</td>
<td>0.001</td>
</tr>
<tr>
<td>(V_{C})</td>
<td>0.33</td>
<td>0.35</td>
<td>0.33</td>
</tr>
<tr>
<td>(H^2)</td>
<td>1.07</td>
<td>1.03</td>
<td>1.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Percent Thin Kernels</th>
<th>Percent Plump Groat</th>
<th>Groat Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
</tr>
<tr>
<td>(V_L)</td>
<td>18.4***</td>
<td>17.9***</td>
<td>18.3***</td>
</tr>
<tr>
<td>(V_G)</td>
<td>25.3***</td>
<td>25.1***</td>
<td>23.0***</td>
</tr>
<tr>
<td>(V_{RI})</td>
<td>2.62***</td>
<td>2.73***</td>
<td>-</td>
</tr>
<tr>
<td>(V_{C})</td>
<td>-</td>
<td>-</td>
<td>1.38***</td>
</tr>
<tr>
<td>(H^2)</td>
<td>0.46</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>RE</td>
<td>-</td>
<td>1.07</td>
<td>1.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Thousand Kernels Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>(V_L)</td>
<td>1.52***</td>
</tr>
<tr>
<td>(V_G)</td>
<td>5.07***</td>
</tr>
<tr>
<td>(V_{RI})</td>
<td>0.27***</td>
</tr>
<tr>
<td>(V_{C})</td>
<td>-</td>
</tr>
<tr>
<td>(V_{C})</td>
<td>5.25***</td>
</tr>
<tr>
<td>(H^2)</td>
<td>0.49</td>
</tr>
<tr>
<td>RE</td>
<td>-</td>
</tr>
</tbody>
</table>

*Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.

M1: most reduced model (Model 1), M2: includes row effect (Model 2), M3: includes column effect (Model 3), and M4: includes both row and column effects (Model 4), \(V_L\): variance due to location, \(V_G\): variance due to genotype, \(V_{RI}\): variance due to row nested in location, \(V_{C}\): variance due to column nested in location, \(H^2\): broad sense heritability, RE: relative efficiency
Table 2.4. Genotypic correlations among milling and nutritional quality traits in 2015 and 2016 (above and below diagonal respectively)

<table>
<thead>
<tr>
<th></th>
<th>BG</th>
<th>FAT</th>
<th>PLUMP</th>
<th>THIN</th>
<th>GP</th>
<th>PPG</th>
<th>TKW</th>
</tr>
</thead>
<tbody>
<tr>
<td>PP</td>
<td>0.14*</td>
<td>0.16*</td>
<td>0.20**</td>
<td>-0.23***</td>
<td>0.30***</td>
<td>0.33***</td>
<td>0.28***</td>
</tr>
<tr>
<td>BG</td>
<td>-0.06</td>
<td>0.04</td>
<td>-0.04</td>
<td>-0.07</td>
<td>-0.05</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>FAT</td>
<td>0.30***</td>
<td>-0.15*</td>
<td>0.13*</td>
<td>0.12</td>
<td>-0.02</td>
<td>-0.08</td>
<td></td>
</tr>
<tr>
<td>PLUMP</td>
<td>-0.01</td>
<td>-0.41***</td>
<td>-0.97***</td>
<td>-0.06</td>
<td>0.81***</td>
<td>0.66***</td>
<td></td>
</tr>
<tr>
<td>THIN</td>
<td>0.11</td>
<td>0.36***</td>
<td>-0.80***</td>
<td>-0.02</td>
<td>-0.82***</td>
<td>-0.67***</td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td>-0.02</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.15*</td>
<td>0.41***</td>
<td>0.18**</td>
<td></td>
</tr>
<tr>
<td>PPG</td>
<td>-0.04</td>
<td>-0.39***</td>
<td>0.83***</td>
<td>-0.82***</td>
<td>0.27***</td>
<td>0.68***</td>
<td></td>
</tr>
<tr>
<td>TKW</td>
<td>0.05</td>
<td>-0.26***</td>
<td>0.71***</td>
<td>-0.68***</td>
<td>0.06</td>
<td>0.66***</td>
<td></td>
</tr>
</tbody>
</table>

*Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.

Table 2.5. Average prediction accuracy for oat milling and nutritional quality traits using cross-validation in each environment for 227 oat breeding lines evaluated in seven environments.

<table>
<thead>
<tr>
<th>Year</th>
<th>Location</th>
<th>PP</th>
<th>BG</th>
<th>FAT</th>
<th>PLUMP</th>
<th>THIN</th>
<th>PPG</th>
<th>GP</th>
<th>TKW</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015</td>
<td>NERF</td>
<td>0.56</td>
<td>0.27</td>
<td>0.66</td>
<td>0.65</td>
<td>0.70</td>
<td>0.62</td>
<td>0.43</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>SERF</td>
<td>0.36</td>
<td>0.38</td>
<td>0.65</td>
<td>0.68</td>
<td>0.63</td>
<td>0.63</td>
<td>0.42</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>Volga</td>
<td>0.54</td>
<td>0.19</td>
<td>0.46</td>
<td>0.65</td>
<td>0.58</td>
<td>0.51</td>
<td>0.47</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Winner</td>
<td>0.39</td>
<td>0.43</td>
<td>0.73</td>
<td>0.74</td>
<td>0.72</td>
<td>0.64</td>
<td>0.48</td>
<td>0.39</td>
</tr>
<tr>
<td>2016</td>
<td>NERF</td>
<td>NA</td>
<td>0.45</td>
<td>0.62</td>
<td>0.44</td>
<td>0.61</td>
<td>0.42</td>
<td>0.60</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>Volga</td>
<td>NA</td>
<td>0.35</td>
<td>0.70</td>
<td>0.48</td>
<td>0.57</td>
<td>0.45</td>
<td>0.63</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Winner</td>
<td>NA</td>
<td>0.30</td>
<td>0.54</td>
<td>0.62</td>
<td>0.52</td>
<td>0.61</td>
<td>0.41</td>
<td>0.39</td>
</tr>
</tbody>
</table>

NA means data not available for that year, PP: percent protein, BG: β-Glucan, FAT: fat content, PLUMP: percent plump kernels, THIN: percent thin kernels, PPG: percent plump groat, GP: groat percent, TKW: thousand kernels weight
Fig. 2. 2. Relationship between heritability and prediction accuracy for oat quality traits measured in the 2015 preliminary yield trial at four South Dakota locations.

\[ y = 0.595x + 0.0404 \]
\[ R^2 = 0.57779 \]

Fig. 2. 3. Relationship between heritability and prediction accuracy for oat quality traits measured in the 2016 preliminary yield trial at three South Dakota locations.

\[ y = 0.3235x + 0.228 \]
\[ R^2 = 0.21016 \]
Fig. 2. 4. Prediction accuracy for percent content using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

Fig. 2. 5. Prediction accuracy for β-glucan concentration using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2.6. Prediction accuracy for fat content using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

Fig. 2.7. Prediction accuracy for percent plump kernels using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2. 6. Prediction accuracy for percent thin kernels using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

Fig. 2. 7. Prediction accuracy for percent plump groat using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2. 8. Prediction accuracy for groat percent using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

Fig. 2. 9. Prediction accuracy for thousand kernels weight using six genomic selection methods for 227 oat genotypes evaluated in the 2015 preliminary yield trials grown at four South Dakota locations.

NS: non-significant, * Significant at 0.05 probability, ** Significant at 0.01 probability and *** Significant at 0.001 probability.
Fig. 2. 10. Prediction accuracy for β-glucan using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

Fig. 2. 11. Prediction accuracy for fat content using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2. 12. Prediction accuracy for percent plump kernels using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

Fig. 2. 13. Prediction accuracy for percent thin kernels using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2. 14. Prediction accuracy for percent plump groat using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

Fig. 2. 15. Prediction accuracy for groat percent using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2.16. Prediction accuracy for thousand kernels weight using six genomic selection methods for 238 oat genotypes evaluated in the 2016 preliminary yield trials grown at three South Dakota locations.

NS: non-significant, *Significant at 0.05 probability, **Significant at 0.01 probability and ***Significant at 0.001 probability.
Fig. 2. 17. Prediction accuracy for β-glucan and fat content using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations.
Fig. 2. 18. Prediction accuracy for groat percent and percent plump groat using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations
Fig. 2. Prediction accuracy for percent plump kernels and percent thin kernels using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations.
Fig. 2. 20. Prediction accuracy for thousand kernels weight using six genomic selection methods for 227 and 238 oat genotypes evaluated in 2015 and 2016 preliminary yield trials grown at different South Dakota locations.
LITERATURE CITED


Crossa, J., de Los Campos, G., Pérez, P., Gianola, D., Burgueño, J., Araus, J. L., Makumbi, D., Singh, R.P., Dreisigacker, S., Yan, J., Arief, V., Banziger, M., &


Wang, Y., Mette, M. F., Miedaner, T., Gottwald, M., Wilde, P., Reif, J. C., & Zhao, Y. (2014). The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC genomics, 15*(1), 556.

