UTILIZING A HISTORICAL WHEAT COLLECTION TO DEVELOP NEW TOOLS FOR MODERN PLANT BREEDING

by

TREVOR W. RIFE

B.S., University of Wyoming, 2011

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

The Green Revolution is credited with saving billions of lives by effectively harnessing new genetic resources and breeding strategies to create high-yielding varieties for countries lacking adequate food security. To keep the next billion people in a state of food security, plant breeders will need to rapidly incorporate novel approaches and technologies into their breeding programs. The work presented here describes new genomic and phenomic strategies and tools aimed at accelerating genetic gain in plant breeding.

Plant breeders have long relied on regional testing networks to evaluate new breeding lines across many locations. These are an attractive resource for both retrospective and contemporary analysis due to the vast amount of data available. To characterize genetic progress of plant breeding programs in the Central Plains, entries from the Southern Regional Performance Nursery dating back to 1992 were evaluated in field trials. The trend for annual improvement was 1.1% yr^{-1}, matching similar reports for genetic gain. During the same time period, growth of on-farm yields stagnated.

Genomic selection, a promising method to increase genetic gain, was tested using historical data from the SRPN. A temporal-based model showed that, on average, yield predictions outperformed a year-to-year phenotypic correlation. A program-based model found that the predictability of a breeding program was similar when using either data from a single program or from the entire regional collection.

Modern DNA marker platforms either characterize a small number of loci or profile an entire genome. Spiked genotyping-by-sequencing (sGBS) was developed to address the need in breeding programs for both targeted loci and whole-genome selection. sGBS uses a low-cost, integrated approach that combines targeted amplicons with reduced representation genotyping-

by-sequencing. This approach was validated using converted and newly-designed markers targeting known polymorphisms in the leaf rust resistance gene *Lr34*.

Plant breeding programs generate vast quantities of data during evaluation and selection of superior genotypes. Many programs still rely on manual, error-prone methods to collect data. To make this process more robust, we have developed several open-source phenotyping apps with simple, intuitive interfaces.

A contemporary Green Revolution will rely on integrating many of these innovative technologies into modern breeding programs.

UTILIZING A HISTORICAL WHEAT COLLECTION TO DEVELOP NEW TOOLS FOR MODERN PLANT BREEDING

by

TREVOR W. RIFE

B.S., University of Wyoming, 2011

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Interdepartmental Genetics
College of Agriculture

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Jesse Poland

Copyright

Abstract

The Green Revolution is credited with saving billions of lives by effectively harnessing new genetic resources and breeding strategies to create high-yielding varieties for countries lacking adequate food security. To keep the next billion people in a state of food security, plant breeders will need to rapidly incorporate novel approaches and technologies into their breeding programs. The work presented here describes new genomic and phenomic strategies and tools aimed at accelerating genetic gain in plant breeding.

Plant breeders have long relied on regional testing networks to evaluate new breeding lines across many locations. These are an attractive resource for both retrospective and contemporary analysis due to the vast amount of data available. To characterize genetic progress of plant breeding programs in the Central Plains, entries from the Southern Regional Performance Nursery dating back to 1992 were evaluated in field trials. The trend for annual improvement was 1.1% yr^{-1}, matching similar reports for genetic gain. During the same time period, growth of on-farm yields stagnated.

Genomic selection, a promising method to increase genetic gain, was tested using historical data from the SRPN. A temporal-based model showed that, on average, yield predictions outperformed a year-to-year phenotypic correlation. A program-based model found that the predictability of a breeding program was similar when using either data from a single program or from the entire regional collection.

Modern DNA marker platforms either characterize a small number of loci or profile an entire genome. Spiked genotyping-by-sequencing (sGBS) was developed to address the need in breeding programs for both targeted loci and whole-genome selection. sGBS uses a low-cost, integrated approach that combines targeted amplicons with reduced representation genotyping-

by-sequencing. This approach was validated using converted and newly-designed markers targeting known polymorphisms in the leaf rust resistance gene *Lr34*.

Plant breeding programs generate vast quantities of data during evaluation and selection of superior genotypes. Many programs still rely on manual, error-prone methods to collect data. To make this process more robust, we have developed several open-source phenotyping apps with simple, intuitive interfaces.

A contemporary Green Revolution will rely on integrating many of these innovative technologies into modern breeding programs.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to acknowledge the invaluable guidance provided by my advisor, Dr. Jesse Poland, without whom this dissertation would never have been completed. Further, my other committee members, Drs. Allan Fritz, Bikram Gill, and Chris Toomajian have all made important intellectual contributions to this work. The technicians and undergrads, including Shuangye Yu, Clayton Seaman, Josh Sharon, Ryan Steeves, Byron Evers, Kathy Kaus, Mitch Keller, Jon Meyer, John Shook, and Annie Eaton, who spent their time working on my projects and their contributions were especially appreciated. During my tenure as a graduate student, I have been fortunate to have made many friends who have profoundly shaped my life and I will cherish all the time we have spent together. Finally, my parents, Charlie and Konne, and my sister, Jenny, have been unwavering in their love and support as I went through the challenge that is graduate school.

Dedication

This work is dedicated to all past, current, and future scientists who endeavor to make the world a better place by dedicating their lives to building and discovering.

Chapter 1 - Advancing Plant Breeding

Measuring progress

The benchmark of progress in plant breeding programs is the creation of lines with novel allele combinations that perform better than their parents. This year-on-year progress is known as genetic gain and is a function of genetic diversity, selection accuracy, selection intensity, and selection cycle time. While most reports of genetic gain in wheat have estimated gain to be ~1% yr^{-1}, some reports have supported the idea that contemporary varieties are approaching a yield plateau (Schmidt, 1984; Graybosch and Peterson, 2010). To evaluate the status of wheat yield gains in the US Central Plains, entries from a regional nursery dating back to 1992 were evaluated for yield and other agronomic traits in a common nursery for three years. Gain within this collection was found to match the common estimate of ~1% yr^{-1}. However, on-farm wheat yields in Kansas during the same time period did not show the same amount of growth, indicating that there is yield gap, due to either a lag in varietal adoption or other confounding factors that are impacting on-farm yield growth in the state.

Improving selection

Genomic selection (GS) is a relatively new technology that makes use of whole-genome markers to predict performance of uncharacterized lines (Meuwissen et al., 2001). GS has the potential to dramatically shorten the plant breeding cycle and increase selection intensity. One major interest in breeding is utilizing historical data for modern genomic selection (Rutkoski et al., 2015). In Chapter 3, the possibility of using historical data in the US Central Plains is evaluated using genotypic and phenotypic data from 1992 forward to create several different GS scenarios and compare predicted values to a phenotypic correlation calculated from lines submitted twice. GS outperformed across-year phenotypic correlation in 13 of 23 years

predicted. In addition, program-by-program predictions were performed using either the entire collection of lines or lines from a single breeding program. Results showed similar predictability using either approach.

Improving genotyping

Rapid advances in next-generation sequencing data output have provided the technology to greatly transform the way we think about plant genomics and breeding. To harness this data for plant breeding and genetics, new approaches that simultaneously discover and type polymorphisms have been developed using reduced representation sequencing (Elshire et al., 2011; Poland and Rife, 2012). These whole-genome profiling approaches have given plant breeders an inexpensive tool that can be used for genetic mapping (Poland et al., 2012a), association studies, and genomic selection (Poland et al., 2012b; Jarquín et al., 2014). Plant breeders, however, also rely on single marker genotyping to select for known loci of importance. Since significant resources and time have been invested in identifying important selection targets, it will be highly valuable to develop a new approach that combines the benefits of whole-genome profiling with the targeted nature of single-marker systems. Chapter 4 describes such an approach and its application to genotype a diverse set of wheat varieties for Lr34.

Mobile phenotyping

Significant efforts are being made to improve the collection of data in the field by implementing high-throughput phenotyping technologies. While the attraction of these new technologies is high, the learning curve and barriers to implementation have led to slow adoption by more-traditional breeders. In addition, much of the technology utilized by these systems is only available at great cost and therefore not as available to breeders in developing countries. As a necessary improvement to more-traditional phenotyping approaches, we have developed

several free mobile apps that promote proper data collection, management, and ontology integration. These apps have been widely adopted in both developed and developing countries, indicating that there is still a large desire for more-traditional tools. Chapter 5 describes Field Book, a note-taking app with more than 1500 users around the world. Appendix C describes additional apps that have been developed to streamline collection, management, and analysis of data being used by plant breeders.

References

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5): e19379.

Graybosch, R.A., and C.J. Peterson. 2010. Genetic Improvement in Winter Wheat Yields in the Great Plains of North America, 1959–2008. Crop Sci. 50(5): 1882.

Jarquín, D., K. Kocak, L. Posadas, K. Hyma, J. Jedlicka, G. Graef, and A. Lorenz. 2014. Genotyping by sequencing for genomic prediction in a soybean breeding population. BMC Genomics 15(1): 740.

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819–29.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2): e32253.

Poland, J.A., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012b. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5(3): 103–113.

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J. 5(3): 92–102.

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, and M.E. Sorrells. 2015. Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. Plant Genome 8(1): 0.

Schmidt, J.W. 1984. Genetic Contributions to Yield Gains in Wheat. p. 89–101. *In* Genetic Contributions to Yield Gains of Five Major Crop Plants. CSSA Special Publication SV - 7. Crop Science Society of America and American Society of Agronomy, Madison, WI.

Chapter 2 - A field-based analysis of genetic improvement in winter wheat yields in the US Central Plains from 1992 to 2014

Abbreviations

SRPN, Southern Regional Performance Nursery; BLUE, best linear unbiased estimate

Abstract

Progress in plant breeding programs is the result of creating and selecting new lines with novel allele combinations that perform better than their parents. This year-on-year improvement is known as genetic gain and is a function of genetic diversity, selection accuracy, selection intensity, and selection cycle time. In order to estimate the gain in wheat breeding in the US Central Plains, lines that were submitted to the collaborative Southern Regional Performance Nursery (SRPN) between 1992 and 2014 were grown in a common nursery for three years at two locations in a single replicate augmented block design. Moderate to high heritability was observed for height (H^2 = 0.88), heading date (H^2 = 0.79), and yield (H^2 = 0.41). From the common growout, genetic gain for yield across the time period was estimated at 1.1% yr^{-1} while individual program genetic gain varied between 0.3% and 2.4% yr^{-1}. Increases in Kansas state on-farm yields during the same time period showed a non-significant trend of 0.13% yr^{-1} and large year-to-year variation. These results suggest that while progress is being made in US Central Plains breeding programs, a yield-gap remains and the same relative progress is not being transferred to on-farm production.

Introduction

Genetic gain, or the year-on-year progress observed in plant breeding, is the benchmark by which plant breeding programs advance and is a function of genetic diversity, selection

accuracy, selection intensity, and selection cycle time. An assessment of the rate of genetic gain within and across breeding programs gives a benchmark for plant breeding as one of the most important tools we have to address food security for a growing world population.

Plant breeding programs must evaluate new breeding lines across many locations to identify the best candidates for release as new varieties. To aid with this evaluation in wheat, collaborative regional testing networks across the US are utilized to characterize line performance. The Hard Winter Wheat Regional Nursery Program was established in 1931 by the US Department of Agriculture (USDA) – Agricultural Research Service (ARS) to measure performance, quality, disease resistance, and other agronomic traits of near-release wheat varieties from breeding programs in the US Midwest. Entries submitted by breeders in the region are evaluated at more than 30 locations along with multiple, common, long-term check cultivars. This nursery has been regularly used to estimate genetic gain over time relative to Kharkof, a tall check variety (Schmidt and Worrall, 1983; Graybosch and Peterson, 2010, 2012).

Previous estimates for genetic gain across the same region have reported varying improvement. Battenfield et al. (2013) provided a good review of global studies measuring genetic gain and also measured modern gain in the Great Plains at 0.40% yr^{-1} relative to the performance of TAM 101, a common check variety. Cox et al. (1988), using 30 varieties that were released throughout the 20^{th} century, found a 1% yr^{-1} increase. Graybosch and Peterson (2010) examined genetic gain for a broad time period (1959-2008) as well as a more narrow period (1984-2008). Gain was reported as 1.1% yr^{-1} increase over Kharkof, the common check variety for the entire time period but this trend was non-significant for the more recent years (Graybosch and Peterson, 2010). Investigating the idea that specific adaptation from individual breeding programs may have led to the modern loss of genetic gain, Graybosch and Peterson

(2012) examined yield gains in predetermined growing regions, again finding a lack of gain in the SRPN, except for where the check variety was poorly adapted. Understandably, this gives some credence to the ideas originally presented by Schmidt (1984) of a slowdown or plateau for genetic gain in recent decades.

However, many of these previous studies have been retrospective and relative, insofar that they use the unbalanced regional nursery data across years and rely on the transformation of the mean entry yield into a relative percent of the yield of a long-term check. This approach assumes minimal genotype-by-environment interaction is present particularly for the long-term check. This assumption is likely not satisfied for the RPN since Kharkof is a tall variety in contrast to all contemporary wheat varieties being semi-dwarf. Kharkof is also better-adapted to cooler environments but is still used for comparison in warm and dry environments of the Southern Plains (Graybosch and Peterson, 2012). Further complicating historical measures of genetic gain, in previous years of this nursery, each participating location maintained their own source of Kharkof, presenting an opportunity for genetic drift and selection, resulting in subsequent phenology and morphological differences (Cox and Worrall, 1987). Other studies have examined genetic gain in wheat in the US using a common nursery experiment, but have evaluated a relatively small number of cultivars (12-35) representing a large number of years (average 3.8 years/entry) (Cox et al., 1988; Donmez et al., 2001; Khalil et al., 2002; Fufa et al., 2005; Battenfield et al., 2013).

To reduce the confounding issues detailed above in assessing the genetic gain of wheat breeding in the U.S. Central Plains, 711 entries that were submitted to the SRPN from 1992-2014 were grown in a common garden for a total of four site-years. Height, heading date, and yield measurements were collected and used to calculate trait heritabilities. Genetic gain from 1992 to

2014 was estimated across the entire collection of entries as well as on a program basis. To determine if this genetic gain was realized in growers' fields, the rate of gain in on-farm yields over the same time period was determined. Increases in Kansas state on-farm yields during the same time period showed a non-significant, slowing trend and large year-to-year variation. Our results suggest that progress in on-farm yields does not match the progress being measured solely from breeding nurseries or that genetic 'gain' in wheat breeding for this region has only been maintenance breeding or the prevention of yield loss due to increasing pathogen pressure and less favorable environments. There are stark implications of reduced gain and this observed yield-gap and the effect it will have on future productivity and food security.

Materials and Methods

Plant material

Seed was acquired from original samples distributed by the SRPN for entries dating back to 1992. Entries were grown in a greenhouse in fall of 2012 to increase the amount of seed and then grown in single rows in the summer of 2013 at Ashland, KS to further increase the amount of seed and allow for replicated testing. For subsequent field trials, 711 entries were chosen on the basis of seed availability and limiting line redundancy (Table 2-1).

Field design and data collection

An augmented block design with two regional check varieties (Everest and TAM 112) was created using the agricolae package in R with ranges corresponding to blocks (de Mendiburu, 2016). Experimental entries were randomly assigned to a block for each environment (location-year). Entries were tested for three years (2014, 2015, and 2016) at two locations (Ashland Bottoms Research Farm near Manhattan, KS and Hays, KS) giving six

location-years of evaluation. In this study, location-years are referenced by the last two digits of the year and first letter of the location (e.g. 14A, 15H).

Entries were evaluated in either 0.75m x 1.22m three row plots (14A, 14H, 15A) or 1.5m x 2.44m six row plots (15H, 16A, 16H). Yield was collected from 14A, 15A, 15H, and 16H with the other two trials being lost due to extreme drought (14H) and flooding (16A). The Android app Field Book was used to collect the following traits: height from 15A and 16H; heading date from 14A and 15A (Rife and Poland, 2014).

Data analysis

Twenty entries that were originally submitted to the SRPN as hybrids were removed from subsequent analysis. Plots that had seed loss or mixing due to harvesting errors were removed from additional analysis (11 in 14A; 10 in 15A). No data was collected or used for analysis from the two trials that were lost (14H and 16A).

Plot-level yields from 14A and 15A were corrected for plot size. Entry yield in each environment was adjusted using the checks within each block. The grand mean of the check varieties in each environment was used to calculate a block adjustment factor, which was used to modify the yield for each entry in the block.

To estimate variance effects, a linear mixed model was created for each trait using the lmer command from the lme4 package in R (Bates et al., 2015). Variance effects were used to calculate heritability with

$$H^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{ge}^2}{e} + \frac{\sigma_{err}^2}{e}} \quad [1]$$

and where σ_g^2 is the genotypic variance, σ_{ge}^2 is the genotype by environment interaction, σ_{err}^2 is the residual error variance, and e is the number of environments (Holland et al., 2003).

Genetic gain

A linear mixed model was fit using the lme4 package in R (Bates et al., 2015) for adjusted yield with

$$y_{ijk} = \mu + g_i + m_j + r_k + (gm)_{ij} + (gr)_{ik} + e_{ijk} \quad [2]$$

where y_{ijk} is the adjusted yield, μ is the overall mean, g_i is the fixed genotype effect for each genotype, m_j is the random effect for each *j*th year with independent and identically distributed (i.i.d.) $m_j \sim N(0, \sigma_j^2)$, r_k is the random effect for each *k*th location with i.i.d. $r_k \sim N(0, \sigma_k^2)$, $(gm)_{ij}$ is the random interaction effect of the *i*th genotype and *j*th year with i.i.d. $(gm)_{ij} \sim N(0, \sigma_{ij}^2)$, $(gr)_{ik}$ is the random interaction effect of the *i*th genotype with the *k*th location with i.i.d. $(gr)_{ik} \sim N(0, \sigma_{ik}^2)$, and e_{ijk} as the random error assumed i.i.d. $e_{ijk} \sim N(0, I\sigma_e^2)$. Best linear unbiased estimates (BLUEs) were extracted from the model using the coef function in R (R Core Team, 2014). The BLUE for each entry was grouped into the year the entry was first evaluated in the SRPN and a linear model was fit with BLUEs as a function of the evaluation year.

Genetic gain within each breeding program was calculated by subsetting the BLUEs by program, and refitting the linear model above. Programs for which fewer than 20 entries were evaluated in this study were excluded from this process.

Genetic gain was also calculated for each location-year by fitting a linear mixed model with adjusted yield as a response, entry as a fixed effect, and submitted year as a random effect. Entry BLUEs were grouped by their evaluation year and a linear model was fit with BLUEs as a function of the evaluation year.

Kansas yield data

Kansas state-wide yield data from 1903 to 2015 was obtained from the USDA National Agricultural Statistics Service (USDA NASS, 2016). Genetic gain over time was measured by

fitting independent linear models with yield (in bushels per acre) for the following time periods: 1903-1960 (pre-Green Revolution; tall wheat), 1961-1980 (Green Revolution; semi-dwarf transition period), 1981-2015 (post-Green Revolution; semi-dwarf wheat), 1992-2014 (years used in this study), and 1960-2014 (modern era semi-dwarf wheat).

Results and Discussion

Phenotypic data

Of the four nurseries that were harvested (14A, 15A, 15H, and 16H), 3,092 plots were planted and 2,991 plots were used in this analysis with a total of 10,911 phenotypic measurements were collected for yield, height, and heading date. Heritability is the ratio of genetically caused variation to the total variation for a given trait (Acquaah, 2007). Across this set of nurseries, moderate to high heritability was observed for height ($H^2 = 0.78$), heading date ($H^2 = 0.79$), and yield ($H^2 = 0.45$). These estimates are in line with similar studies in the same region (Häberle et al., 2007; Zhang et al., 2015).

A barrier to progress in breeding programs is inaccurate and incomplete data collection. Even in this limited experiment, 1/3 of the planted locations were lost, demonstrating how difficult plant breeding can be in a region with large environmental variance.

Genetic gain

Measuring genetic gain is useful to understand the amount of progress that has been made in plant breeding programs. Genetic gain from this collection of entries was an estimated 1.1% yr^{-1} (95% CI 0.9 – 1.29%) (Figure 2-1). Comparatively, this figure is higher than other measures of genetic gain in studies that have examined similar time periods (Graybosch and Peterson, 2010, 2012). Substantial variability was observed for yield within each grouping year.

While the calculated gain was higher than other studies, it's difficult to determine if this gain was due to improved genetics or depressed yields for older varieties. So-called 'maintenance breeding' that keeps the most recent variety yield at a certain threshold in response to recent biotic and abiotic stresses could potentially be responsible for the observed gain. Supporting this idea, 2016, the evaluation year that showed the largest genetic gain among the lines evaluated also experienced increased biotic pressure from stripe rust and leaf rust contributing to yield loss in older varieties.

To evaluate the progress that has been made within each program, entries were subsetted based on program and gain was recalculated. There was substantial variation of gain across different breeding programs with gain within individual programs ranging from 0.37 to 1.92% yr^{-1} (Figure 2-4, Table 2-2). Due to the relatively few number of lines to represent each program, there was large error around percent gain estimates were observed in nearly every program.

Multiple breeding targets may also be responsible for some of the variation observed between breeding programs. For instance, Oklahoma State University selects wheat varieties for high grain yield but also focuses on developing wheat varieties that produce substantial winter forage, often with a yield tradeoff. The combination of breeding lines from the same program but with contrasting breeding targets creates the possibility of limiting genetic gain for the breeding program as a whole in the yield target environments evaluated for this study.

On-farm yields

Although the ultimate goal for a breeder is to create lines that perform significantly better than his or her own experimental lines, the success of a new variety and the plant breeding enterprise as a whole, is the transfer of these genetic gains to increased farm yield. To evaluate to what extent genetic gain has been transferred from breeding programs to farmers, data from

Kansas state-wide yields from 1903 through 2015 was used to determine the average yield increase per year over several different time periods. Time periods were chosen based on years with similar agronomic practices. The five different time periods for which yield gain was calculated included 1903-1960, corresponding to tall wheat varieties with less-intense wheat breeding and agronomic management; 1961-1980, corresponding to the introduction of semi-dwarf wheat varieties and increased nitrogen application; 1981-2015, corresponding to contemporary breeding and complete adoption of semi-dwarf varieties; 1960-2015, corresponding to modern breeding and semi-dwarf varieties; and 1992-2014, corresponding to the same years used in this study. A linear model was used to estimate the amount of gain in each time period (Figure 2-2).

There were substantially different yield gains during these time periods, roughly corresponding to the implementation and exploitation of different agronomic and genetic technologies. The period relating to increasing adoption of semi-dwarf varieties and more applied nitrogen had substantially more gain than any other evaluated time period. Dividing the time period since the introduction of semi-dwarf varieties into several intervals indicated that yield gains on-farm are decelerating. This could be due to a number of factors including disease pressure, environmental stresses (Lobell et al., 2011), implementation of agricultural practices such as no-till, agricultural intensification, or factors related to a changing climate.

Of interest is the fact that when evaluating modern wheat varieties as a single time period (i.e. 1960-2015), the rate of gain matches the generally accepted 1% yr^{-1} (Tester and Langridge, 2010; Battenfield et al., 2013). However, this estimate is innately due to the inclusion of the extreme growth of on-farm yields from 1960-1980 and is therefore misleading.

The time period corresponding to the years used in this study had the least amount of yield gain at only 0.13%.

Conclusions

Much work has been done to quantify the current rate of genetic gain as well as the required rate to sustain current trends in population growth and meet projected food demand (Tester and Langridge, 2010). The current accepted perspective is that we must significantly increase progress relative to the historical rate of gain, and in many situations need to double the rate of gain (Tester and Langridge, 2010; Ray et al., 2013). However, this conclusion is based on the idea that the current rate of genetic gain is a continuation of the significant increases that were seen during the Green Revolution, which is clearly not the case. While we found gain to be positive in both experimental and on-farm environments, the proportion increase seen in the on-farm yields during the time period of interest highlights the fact that gains in breeding productivity may not be making their way back to the most important stakeholders. In contrast to the Green Revolution when the gains were substantial, the nominal gains experienced during contemporary breeding have been slower to move back to farmers. New genetic or technological enhancements will be necessary to improve this current trend.

Acknowledgements

The USDA-NIFA funded Triticeae Coordinated Agriculture Project (TCAP) (2011-68002-30029) provided support for T. Rife. This work was completed under the auspices of WGRC I/UCRC partially funded by NSF grant contract (IIP-1338897) and the USAID Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051). Mention of trade names does not constitute endorsement by the U.S. Department of Agriculture.

References

Acquaah, G. 2007. Principles of Plant Genetics and Breeding: Second Edition.

Bates, D., M. Maechler, B.M. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models using {lme4}. J. Stat. Softw. 67(1): 1–48.

Battenfield, S.D., A.R. Klatt, and W.R. Raun. 2013. Genetic Yield Potential Improvement of Semidwarf Winter Wheat in the Great Plains. Crop Sci. 53(3): 946.

Cox, T., J. Shroyer, and L. Ben-Hui. 1988. Genetic improvement in agronomic traits of hard red winter wheat cultivars 1919 to 1987. Crop Sci. 28(5).

Cox, T., and W. Worrall. 1987. Electrophoretic variation among and within strains of "Kharkof"wheat maintained at 11 locations. Euphytica 36: 815–822.

Donmez, E., R.G. Sears, J.P. Shroyer, and G.M. Paulsen. 2001. Genetic gain in yield attributes of winter wheat in the Great Plains. Crop Sci. 41(5): 1412–1419.

Fufa, H., P.S. Baenziger, B.S. Beecher, R. a. Graybosch, K.M. Eskridge, and L. a. Nelson. 2005. Genetic improvement trends in agronomic performances and end-use quality characteristics among hard red winter wheat cultivars in Nebraska. Euphytica 144(1–2): 187–198.

Graybosch, R.A., and C.J. Peterson. 2010. Genetic Improvement in Winter Wheat Yields in the Great Plains of North America, 1959–2008. Crop Sci. 50(5): 1882.

Graybosch, R.A., and C.J. Peterson. 2012. Specific Adaptation and Genetic Progress for Grain Yield in Great Plains Hard Winter Wheats from 1987 to 2010. Crop Sci. 52(2): 631.

Häberle, J., M. Schmolke, G. Schweizer, V. Korzun, E. Ebmeyer, G. Zimmermann, and L. Hartl. 2007. Effects of two major fusarium head blight resistance QTL verified in a winter wheat backcross population. Crop Sci. 47(5): 1823–1831.

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martínez. 2003. Estimating and Interpreting Heritability for Plant Breeding: An Update. p. 9–112. *In* Plant Breeding Reviews. John Wiley & Sons, Inc., Oxford, UK.

Khalil, I.H., B.F. Carver, E.G. Krenzer, C.T. MacKown, and G.W. Horn. 2002. Genetic trends in winter wheat yield and test weight under dual-purpose and grain-only management systems. Crop Sci. 42(3): 710–715.

Lobell, D.B., W. Schlenker, and J. Costa-Roberts. 2011. Climate trends and global crop production since 1980. Science (80-.). 333(6042): 616–620.

de Mendiburu, F. 2016. agricolae: Statistical Procedures for Agricultural Research.

R Core Team. 2014. R: A language and environment for statistical computing.

Ray, D.K., N.D. Mueller, P.C. West, and J. a Foley. 2013. Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLoS One 8(6): e66428.

Rife, T.W., and J.A. Poland. 2014. Field Book: An Open-Source Application for Field Data Collection on Android. Crop Sci. 54(4): 1624–1627.

Schmidt, J.W. 1984. Genetic Contributions to Yield Gains in Wheat. p. 89–101. *In* Genetic Contributions to Yield Gains of Five Major Crop Plants. CSSA Special Publication SV - 7. Crop Science Society of America and American Society of Agronomy, Madison, WI.

Schmidt, J., and W. Worrall. 1983. Trends in yield improvement through genetic gains. p. 691–700. *In* Proceedings of the 6th International Wheat Genetics Symposium. Kyoto, Japan.

Tester, M., and P. Langridge. 2010. Breeding Technologies to Increase Crop Production in a Changing World. Science (80-.). 327(5967): 818–822.

USDA NASS. 2016. NASS - Kansas. Available at https://www.nass.usda.gov/Statistics_by_State/Kansas/index.php (verified 1 August 2016).

Zhang, G., R. Aiken, and T.J. Martin. 2015. Relationship between carbon isotope discrimination and grain yield of rainfed winter wheat in a semi-arid region. Euphytica 204(1): 39–48.

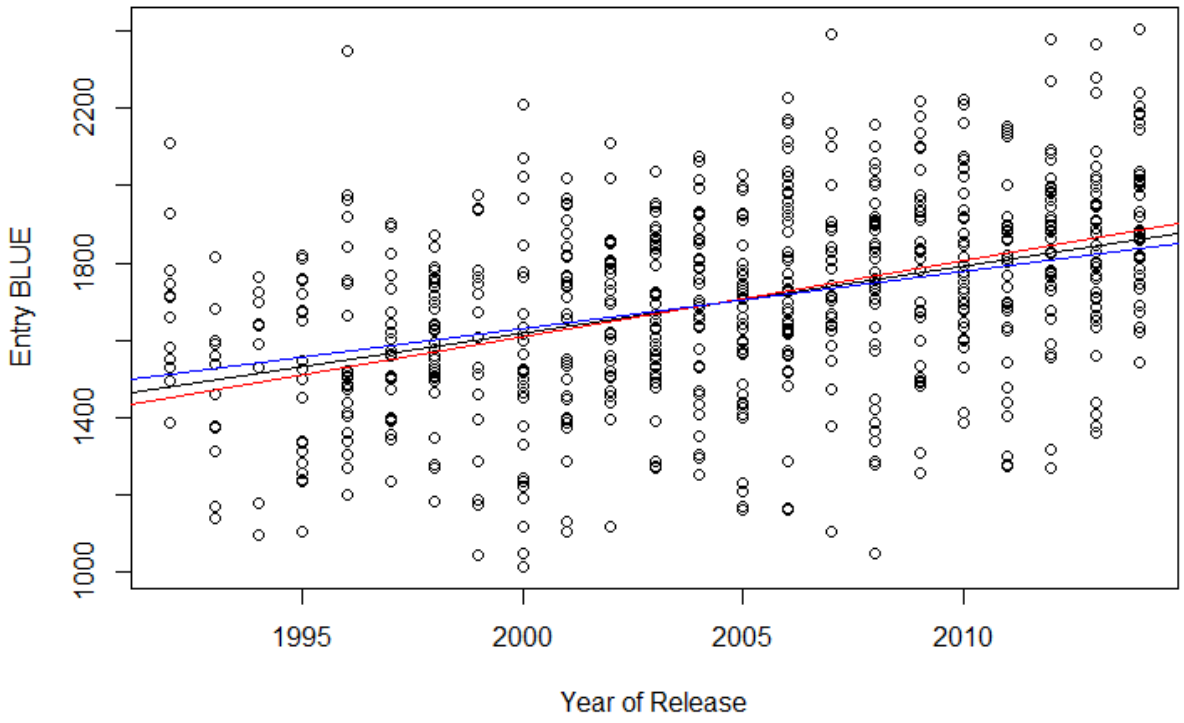

Figure 2-1. Entry BLUEs plotted against the year they were evaluated in the SRPN. Black line indicates linear regression of Entry BLUE on Year of Release (slope = 17.25). Red and blue lines indicate 95% confidence interval around the regression line.

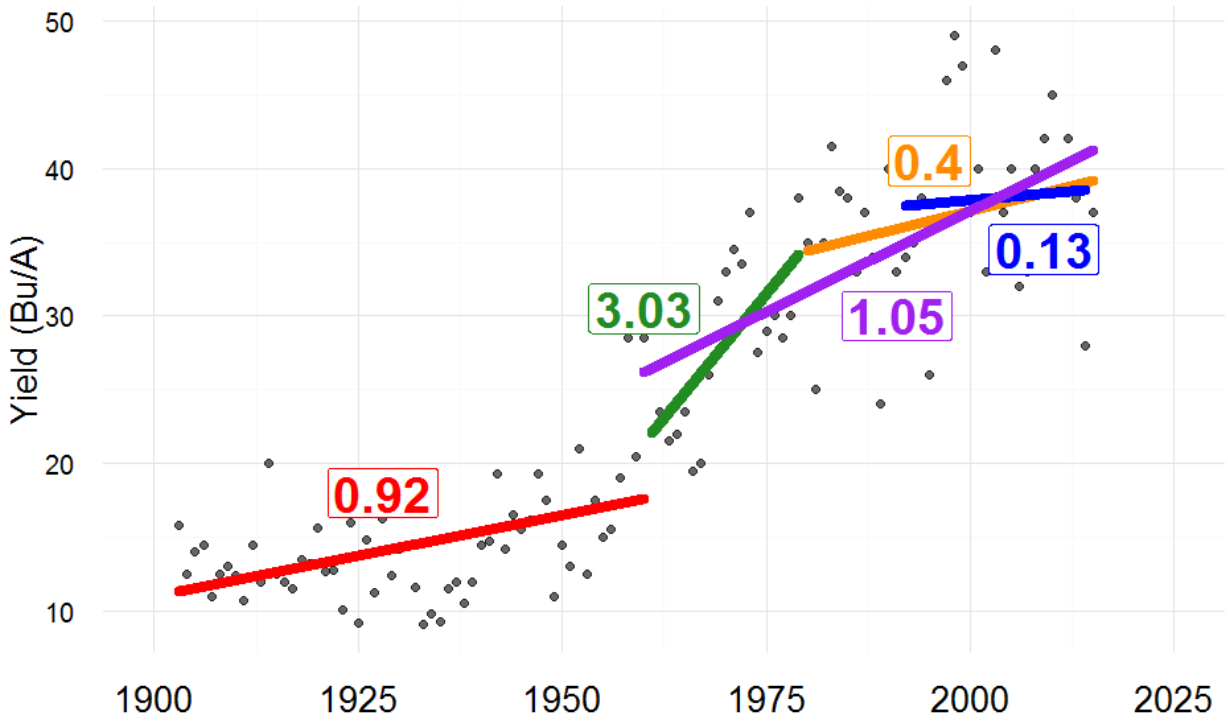

Figure 2-2. On-farm yield trends and genetic gain in percent gain per year for 1903-1960 (red), 1961-1980 (green), 1981-2015 (orange), 1961-2015 (purple), and 1992-2014 (blue).

Table 2-1. The number of entries used in this study grouped by their original year they were evaluated. In total, 711 entries were evaluated.

Year	Count
1992	14
1993	15
1994	12
1995	24
1996	24
1997	31
1998	35
1999	18
2000	31
2001	32
2002	29
2003	38
2004	37
2005	41
2006	42
2007	29
2008	36
2009	33
2010	33
2011	33
2012	44
2013	40
2014	40

Table 2-2. Genetic gain for each program that submitted more than twenty lines to the SRPN between 1992 and 2014.

Program	Year	Lines	Percent Gain	Lower	Upper
AgriPro	1992	65	1.66	0.99	2.43
Colorado State University	1994	75	0.90	0.37	1.51
Kansas State University	1992	66	1.02	0.66	1.41
Kansas State University - Hays	1992	36	1.36	0.72	2.12
Monsanto	1996	61	1.38	0.85	1.97
Oklahoma State University	1992	94	0.81	0.45	1.20
Texas A&M University - Amarillo	1992	43	0.37	-0.31	1.24
Texas A&M University - Vernon	1993	42	1.39	0.62	2.29
Trio Research Inc.	1992	50	1.64	0.82	2.60
University of Nebraska	1992	65	1.92	1.30	2.64

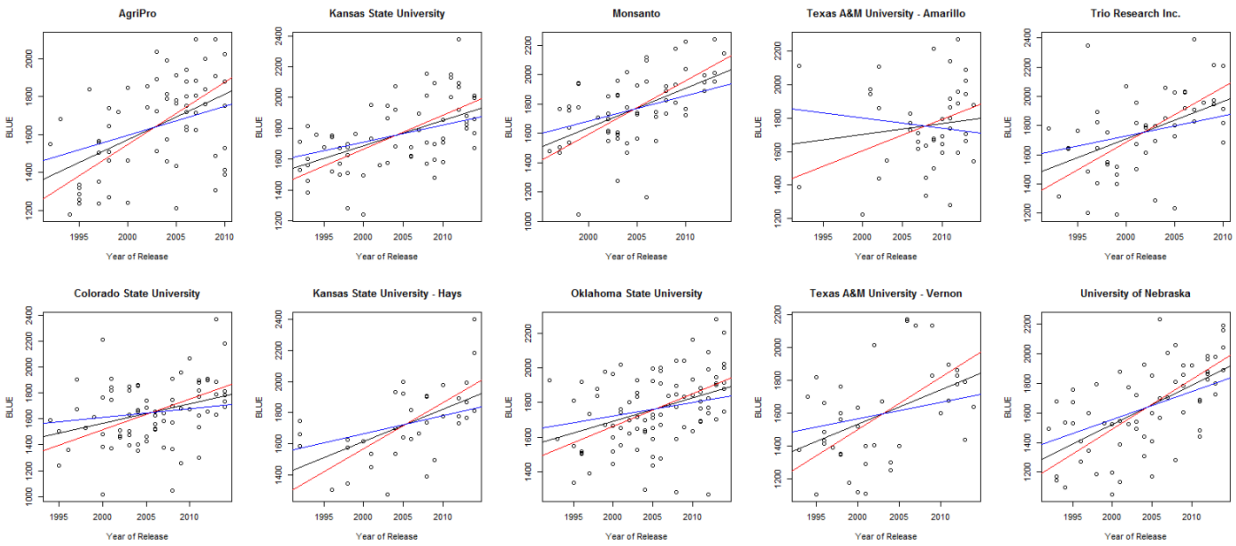

Figure 2-3. Entry BLUEs plotted against the year they were evaluated in the SRPN. Each panel represents a different breeding program. In each panel, the black line indicates the linear regression of the Entry BLUEs for the given breeding program on the Year of Release. The red and blue lines indicate 95% confidence interval around each regression line.

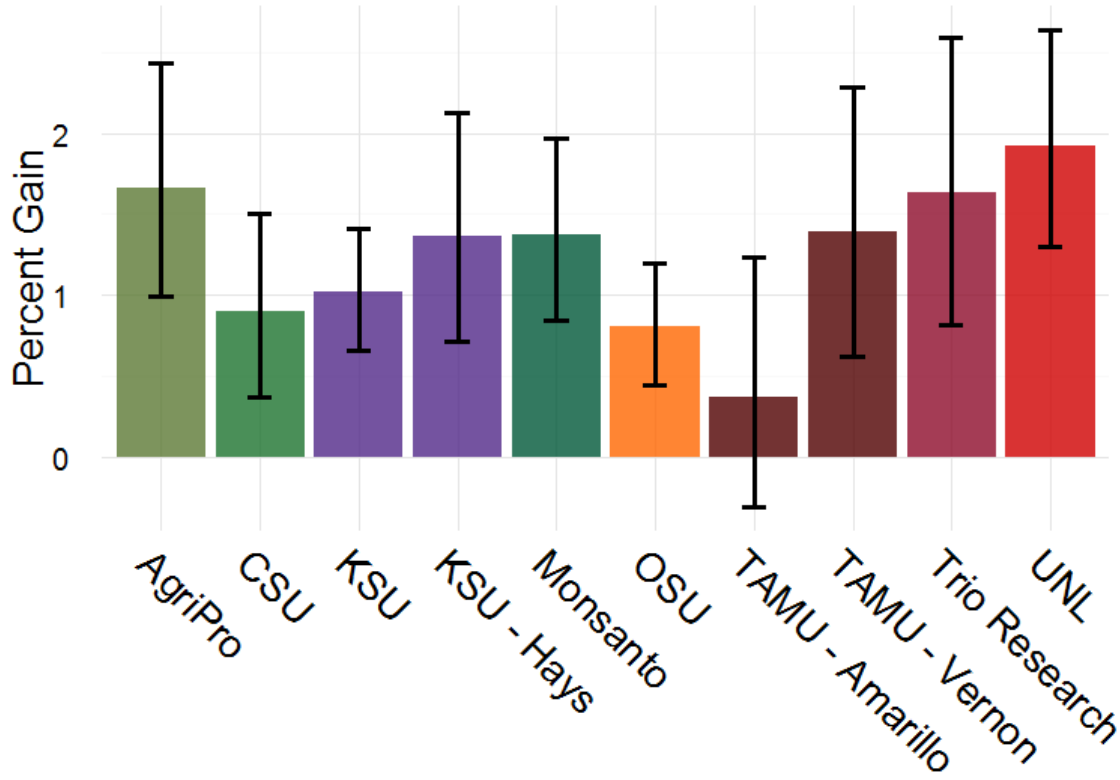

Figure 2-4. Percent gain by program with 95% confidence intervals.

Chapter 3 - Genomic analysis and prediction within a US public collaborative winter wheat regional testing nursery

This chapter is to be submitted to The Plant Genome as the following article:

Rife, T.W., R.A. Graybosch, J.A. Poland. 2016. Genomic analysis and prediction within a US public collaborative winter wheat regional testing nursery.

Abbreviations

SRPN, Southern Regional Performance Nursery; GS, genomic selection; TP, training population; BP, breeding population; BLUP, best linear unbiased predictor

Abstract

The development of inexpensive, whole-genome profiles enables transition to allele-based breeding using genomic prediction models which take into account alleles shared between lines to predict phenotypes and select new lines based on estimated breeding values. This approach can leverage highly-unbalanced datasets common to breeding programs. The Southern Regional Performance Nursery (SRPN) is a public nursery established by the USDA-ARS in 1931 to characterize performance and quality of near-release wheat varieties from breeding programs in the US Central Plains. New entries are submitted annually and can be reentered only once. The trial is grown at more than 30 locations each year and lines are evaluated for grain yield, disease resistance, and agronomic traits. Overall genetic gain is measured across years by including common check cultivars for comparison. We have generated whole-genome profiles via genotyping-by-sequencing for 939 SPRN entries dating back to 1992. We measured the diversity within the nursery and have explored its potential use as a GS training population. GS prediction models for yield across years (average r= 0.33) outperformed observed phenotypic

correlation across years (r=0.27) for a majority of the years evaluated, suggesting that genomic selection has the potential to outperform low heritability selection on yield in these highly variable environments. We also examined the predictability of programs using both program-specific and whole-set training populations. Generally, the predictability of a program was similar with both approaches. These results suggest that wheat breeding programs can collaboratively leverage shared data and provide breeders a means to employ the immense datasets that are generated from regional testing networks.

Introduction

Plant breeding programs exert considerable effort evaluating new breeding lines across many locations to identify superior-performing candidates for release as new varieties. For this evaluation in wheat, collaborative regional testing networks have been developed in the U.S. to provide additional information to breeders on the broad performance of their lines.

The cooperative regional performance testing program was established in 1931 by the USDA-ARS in partnership with university agricultural experiment stations to characterize performance, quality, disease resistance, and other agronomic traits of near-release wheat varieties from breeding programs in the US Central Plains. In this network, the Southern- and Northern Regional Performance Nurseries (SRPN and NRPN) were established where breeders submit entries that are distributed for evaluation at more than 30 locations along with multiple, common, long-term check cultivars (Figure 3-1). Phenotypic data collected from the nurseries includes grain yield, test weight, plant height, lodging, and resistance to a variety of diseases. The regional performance nurseries have been used to regularly measure genetic gain over time (Schmidt and Worrall, 1984; Graybosch and Peterson, 2010, 2012), evaluate long-term wheat

diversity (Cox and Worrall, 1987), and cluster experimental locations into production zones based on performance data (Peterson, 1992).

Previous investigation of broad genotypic characteristics of the RPNs has been limited due to the overall number of lines that have been tested, difficulty in obtaining a complete set of evaluated entries, and an inherent challenge in generating a sufficient amount of genotypic data for each entry. With the recent development of inexpensive, high-density genetic markers, whole-genome marker profiles can now be obtained for every experimental line, making possible new analyses that rely on large amounts of genomic data including diversity studies and genomic selection (Poland and Rife, 2012).

Genomic selection (GS) is a statistical approach that is used to predict phenotypes and select new lines in breeding programs based on favorable allelic combinations (Meuwissen et al., 2001). Breeding programs are investigating and utilizing GS as a tool to shorten the breeding cycle (Heffner et al., 2009, 2010) and increase the selection intensity (Cros et al., 2015; Battenfield et al., 2016). GS has two fundamental components: 1) a population that has been both phenotyped and genotyped which is used to train the prediction model and 2) a population that has been only genotyped to which the model is then applied. Previous literature has assigned each of these two populations various designations (Rincent et al., 2012; Isidro et al., 2014; Rutkoski et al., 2015). Here we will refer to the two populations as the training population (TP) and the breeding population (BP), respectively.

Optimal design of the TP is a research topic of high interest to the breeding community as the phenotypic evaluation of the training population remains a time-consuming and expensive endeavor (Isidro et al., 2014; Akdemir et al., 2015; Spindel et al., 2015). The characteristics that make up an ideal training population are still relatively poorly understood. However, two

features have been promoted as compelling factors: size and degree of relatedness. A correlation exists between the number of lines used in the training population and the accuracy of the predictions (Zhong et al., 2009). However, there are diminishing returns (Asoro et al., 2011). It is not possible to estimate allele effects if there are no common alleles and lack of relationship between the TP and BP. A TP that is more closely related to the BP often results in better prediction accuracy (Hayes et al., 2009; Long et al., 2011; Pszczola et al., 2012; Rutkoski et al., 2015).

The broad scope and design of the RPN makes it an ideal collection to investigate both of these factors since thousands of lines have been evaluated in this nursery. The simultaneous interrelation and stratification of alleles between the regional breeding programs makes it possible to examine how relatedness factors into accuracy both across and within the program.

A successful implementation of GS using the lines that have been evaluated in the RPN would allow plant breeders in the region to leverage this data to transition to allele-based breeding and for predicting stable broad adaptation. Prediction models that take into account alleles shared between lines would make it possible to utilize the vast quantities of phenotypic data available from this nursery. To this end, we have generated whole-genome profiles via genotyping-by-sequencing for SRPN entries dating back to 1992. This genetic data was used to examine SRPN diversity, characterize the potential for this collection to serve as a TP for GS, and evaluate prediction differences between breeding programs.

Materials and Methods

Plant material

A collection of 939 entries (691 unique lines) that were submitted to the Southern Regional Performance Nursery (SRPN) between 1992-2012 was assembled and DNA was

extracted from seedling leaf tissue using a BioSprint 96 DNA Plant Kit (Qiagen). DNA was quantified in plates using PicoGreen and normalized to 20µg/µL (Figure 3-2).

Library construction and data processing

Fourteen GBS libraries were prepared following the protocol detailed by Poland et al. (2012). Briefly, DNA was digested with PstI and MspI and barcoded adapters were ligated to the ends of the fragments. Samples were then pooled at 192-plex, amplified, and sequenced on an Illumina HiSeq 2000. SNPs were called using the approach of Poland et al. (2012b) using a population-based filter. SNPs were filtered to have at least a 5% minor allele frequency and at least 20% data present. For subsequent genomic prediction, entries for which genotypic data was unavailable but had been evaluated in the SRPN in a different year (and as a different entry) were "imputed" if genotypic data was available from a different entry.

Diversity analysis

Check entries that are unrepresentative of current wheat cultivars as well as hybrid varieties for which original seed was not available were removed from subsequent analysis. The SNP calls from the remaining 889 entries (665 unique lines) were used to measure diversity using Nei's genetic distance (Nei, 1973) across the entire collection and on an individual breeding program level with a custom script in R (R Core Team, 2014). For lines that were submitted to the SRPN more than once, only the first entry was used when computing genetic distance. Programs contributing fewer than ten entries (ARS-Manhattan, South Dakota State University, Trigen, and Bayer CropScience) were excluded from the analysis to ensure a less-biased estimate.

Phenotypic data

Historical phenotypic data from 82,546 plots was compiled and a mixed linear model was used to calculate best linear unbiased predictors (BLUPs) for lines with random effects for entry, a random effect for location, year, location by year, and replication within location by year using the lmer command from the lme4 package in R (Bates et al., 2015). The SRPN allows lines to be submitted to the nursery twice, generally for two consecutive years. As a comparison for the genomic prediction accuracies, a phenotypic correlation for yield across years was calculated using 207 entries that were submitted to the nursery for two consecutive years.

Genomic selection

A realized additive relationship matrix (A) was constructed using the A.mat function in the rrBLUP package in R (Endelman, 2011). Markers were imputed using the EM algorithm and a maximum missing threshold of 0.8 was used. The kin.blup function in the rrBLUP package was then used to perform genomic prediction with K set to A (Endelman, 2011). Two separate TP schemes were evaluated. The first was a temporal-based TP constructed such that all lines tested in previous years were used as the TP for a given year resulting in a TP that increased in size for each subsequent cycle. After running the predictions for all years, one significant outlier year (2001) was excluded from the training population and predictions were performed again.

The second approach examined the prediction accuracy of lines from a given breeding program using a TP consisting of either a) all lines from all of the programs or b) other lines from the same program. The former method was performed using a "leave one out" prediction across all entries, subsetting the predicted values by breeding program, and then calculating a correlation between the predicted values and generated BLUPs. The latter method used a "leave one out" approach within the entries from each breeding program.

Results and Discussion

Genotyping

To move from line-based breeding to allele-based breeding methods, a whole-genome profile is needed to calculate a realized relationship matrix. In this study, we utilized genotyping-by-sequencing to produce our genetic data. Using an internal alignment-based pipeline, 53,672 SNPs were discovered and typed with 2,463 of these SNPs having more than 80% data present.

Diversity analysis

Genetic diversity is an important factor for maintaining long term gain in plant breeding programs. To assess the genetic diversity within and across programs, we calculated Nei's genetic distance. Nei's genetic distance for the entire collection of lines was 0.264 with the level of diversity within individual programs ranging from 0.198 to 0.25 (Table 3-1). There was considerable variation in the diversity estimates between breeding programs. As might be expected, there was a linear relationship between the number of lines submitted by a program and its measured level of diversity ($r2 = 0.485$, $p=0.05$). However, there were several exceptions where programs with large numbers of submitted lines had a relatively low measured diversity, potentially due to a narrow breeding program either in target region or germplasm base.

Phenotypic data analysis

Yield data from 82,546 plots, representing 670 unique location-year nurseries, was used in a mixed linear model to calculate a BLUP for each entry. The majority of entries submitted to the SRPN are only tested for a single year, making absolute yield comparisons across all years impossible. However, since 207 of the lines submitted to the SRPN were evaluated in the nursery twice, it was possible to use the performances (i.e. BLUP) of these lines from their first year and second year in the nursery to estimate the phenotypic correlation expected in the nursery. The

correlation for plot yield across years in these lines was moderately low at 0.27 ($p<0.01$). This is to be expected due to the wide range of environments from which data is being generated and the high year-to-year variation common to the Central Plains.

Genomic prediction across years

A temporal-based training population was created that used data from all previous years to make predictions on the next year. Genomic prediction using this approach resulted in an average correlation between the calculated BLUPs and predicted values of 0.33 (Figure 3-3). The correlations for eleven of the predicted years were significant at $p<0.05$.

This approach created a training population that increased in size with each subsequent prediction cycle. However, there was not an observed positive trend in prediction accuracy with the increased training population size. The likely cause is the large influence that the year of evaluation has on the yield of entries within the nursery (Dawson et al., 2013; Lado et al., 2016).

For GS to be implemented into breeding programs, it needs to have similar accuracy to or surpass the selection methods being used by breeders, namely, phenotypic selection. To put our GS predictions into a phenotypic context, we compared them to the phenotypic correlation of lines that were evaluated multiple times in the SRPN. Predictions were superior to the phenotypic correlation in 12 of the 23 years predicted and within the 95% confidence interval of the phenotypic correlation in all except two years (Figure 3-3). One potential explanation for the drastic decrease in predictive accuracy in 2001 is an epidemic of stripe rust (Line, 2002).

Genomic prediction across breeding programs

To determine if data from other breeding programs can be used for genomic prediction within a given breeding program, separate training populations consisting of all experimental lines (excluding the line being predicted) and lines specific to a given breeding program

(excluding the line being predicted) were used to predict lines one at a time within a breeding program (Figure 3-4).

There is a trend in prediction accuracy that is independent of the approach utilized. Breeding programs that are relatively 'predictable' with one method are also relatively 'predictable' with the other. This implies that the potential for a breeding program to implement genomic selection is likely to be founded on characteristics intrinsic to a given program and that, as tested here, genomic selection may not be a suitable selection approach for all breeding programs.

Conclusions

Maintaining long-term, regional testing networks, as well as their seed stocks, can provide additional information for genetic improvement and ensure future crop production and food security. The potential to use existing datasets for new breeding approaches, like genomic selection, is attractive since generating new phenotypes is both cost- and time-prohibitive and the sampling of many past years of environments is invaluable. In this study, we examined the diversity of the Southern Regional Performance Nursery and considered multiple approaches to implement genomic selection using historical data. Genomic predictions across the entire collection outperformed a year-to-year phenotypic correlation (i.e. phenotypic selection accuracy). However, these results were not consistent across breeding programs with several programs showing reduced or no predictive ability. Our results indicate that there may be inherent characteristics of breeding programs such as germplasm base or target region that prohibit or constrain the use of information from other breeding programs and regional testing networks for genomic prediction as a tool for selection. With the increasing need to maximize genetic gain and accelerate delivery of improved high-yielding varieties, the use of historical

data from coordinated testing networks can be a valuable addition to the genomic prediction models used by plant breeders.

Acknowledgements

The USDA-NIFA funded Triticeae Coordinated Agriculture Project (T-CAP) (2011-68002-30029) provided support for T. Rife. This work was completed under the auspices of WGRC I/UCRC partially funded by NSF grant contract (IIP-1338897) and the USAID Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051). Mention of trade names does not constitute endorsement by the U.S. Department of Agriculture.

References

Akdemir, D., J.I. Sanchez, and J.-L. Jannink. 2015. Optimization of genomic selection training populations with a genetic algorithm. Genet. Sel. Evol. 47(1): 38.

Asoro, F.G., M. a. Newell, W.D. Beavis, M.P. Scott, and J.-L. Jannink. 2011. Accuracy and Training Population Design for Genomic Selection on Quantitative Traits in Elite North American Oats. Plant Genome 4(2): 132.

Bates, D., M. Maechler, B.M. Bolker, and S. Walker. 2015. Fitting Linear Mixed-Effects Models using {lme4}. J. Stat. Softw. 67(1): 1–48.

Battenfield, S.D., C. Guzmán, R.C. Gaynor, R.P. Singh, R.J. Peña, S. Dreisigacker, A.K. Fritz, and J.A. Poland. 2016. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. Plant Genome 0(0): 0.

Cox, T., and W. Worrall. 1987. Electrophoretic variation among and within strains of "Kharkof"wheat maintained at 11 locations. Euphytica 36: 815–822.

Cros, D., M. Denis, J.-M. Bouvet, and L. Sánchez. 2015. Long-term genomic selection for heterosis without dominance in multiplicative traits: case study of bunch production in oil palm. BMC Genomics 16(1): 651.

Endelman, J.B. 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4(3): 250.

Graybosch, R.A., and C.J. Peterson. 2010. Genetic Improvement in Winter Wheat Yields in the Great Plains of North America, 1959–2008. Crop Sci. 50(5): 1882.

Graybosch, R.A., and C.J. Peterson. 2012. Specific Adaptation and Genetic Progress for Grain Yield in Great Plains Hard Winter Wheats from 1987 to 2010. Crop Sci. 52(2): 631.

Hayes, B.J., P.M. Visscher, and M.E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. Genet. Res. (Camb). 91(1): 47–60.

Heffner, E.L., A.J. Lorenz, J.-L. Jannink, and M.E. Sorrells. 2010. Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. Crop Sci. 50(5): 1681.

Heffner, E.L., M.E. Sorrells, and J.-L. Jannink. 2009. Genomic Selection for Crop Improvement. Crop Sci. 49(1): 1.

Isidro, J., J.-L. Jannink, D. Akdemir, J. Poland, N. Heslot, and M.E. Sorrells. 2014. Training set optimization under population structure in genomic selection. Theor. Appl. Genet.

Line, R.F. 2002. Stripe Rust of Wheat and Barley in North America: A Retrospective Historical Review. Annu. Rev. Phytopathol. 40(1): 75–118.

Long, N., D. Gianola, G.J.M. Rosa, and K.A. Weigel. 2011. Long-term impacts of genome-enabled selection. J. Appl. Genet. 52(4): 467–480.

Meuwissen, T.H., B.J. Hayes, and M.E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157(4): 1819–29.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U. S. A. 70(12): 3321–3.

Peterson, J.C. 1992. Similarities among Test Sites Based on Cultivar Performance in the Hard Red Winter Wheat Region. Crop Sci. 32(4): 907.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2): e32253.

Poland, J.A., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012b. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5(3): 103–113.

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J. 5(3): 92–102.

Pszczola, M., T. Strabel, H. a Mulder, and M.P.L. Calus. 2012. Reliability of direct genomic values for animals with different relationships within and to the reference population. J. Dairy Sci. 95(1): 389–400.

R Core Team. 2014. R: A language and environment for statistical computing.

Rincent, R., D. Laloë, S. Nicolas, T. Altmann, D. Brunel, P. Revilla, V.M. Rodríguez, J.

Moreno-Gonzalez, a Melchinger, E. Bauer, C.-C. Schoen, N. Meyer, C. Giauffret, C. Bauland, P. Jamin, J. Laborde, H. Monod, P. Flament, a Charcosset, and L. Moreau. 2012. Maximizing the reliability of genomic selection by optimizing the calibration set of reference individuals: comparison of methods in two diverse groups of maize inbreds (Zea mays L.). Genetics 192(2): 715–28.

Rutkoski, J., R.P. Singh, J. Huerta-Espino, S. Bhavani, J. Poland, J.L. Jannink, and M.E. Sorrells. 2015. Efficient Use of Historical Data for Genomic Selection: A Case Study of Stem Rust Resistance in Wheat. Plant Genome 8(1): 0.

Schmidt, J., and W. Worrall. 1984. Trends in yield improvement through genetic gains. p. 691–700. *In* Proceedings of the 6th International Wheat Genetics Symposium.

Spindel, J., H. Begum, D. Akdemir, P. Virk, B. Collard, E. Redoña, G. Atlin, J.-L. Jannink, and S.R. McCouch. 2015. Genomic Selection and Association Mapping in Rice (Oryza sativa): Effect of Trait Genetic Architecture, Training Population Composition, Marker Number and Statistical Model on Accuracy of Rice Genomic Selection in Elite, Tropical Rice Breeding Lines (R Mauricio, Ed.). PLOS Genet. 11(2): e1004982.

Zhong, S., J.C.M. Dekkers, R.L. Fernando, and J.-L. Jannink. 2009. Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a Barley case study. Genetics 182(1): 355–64.

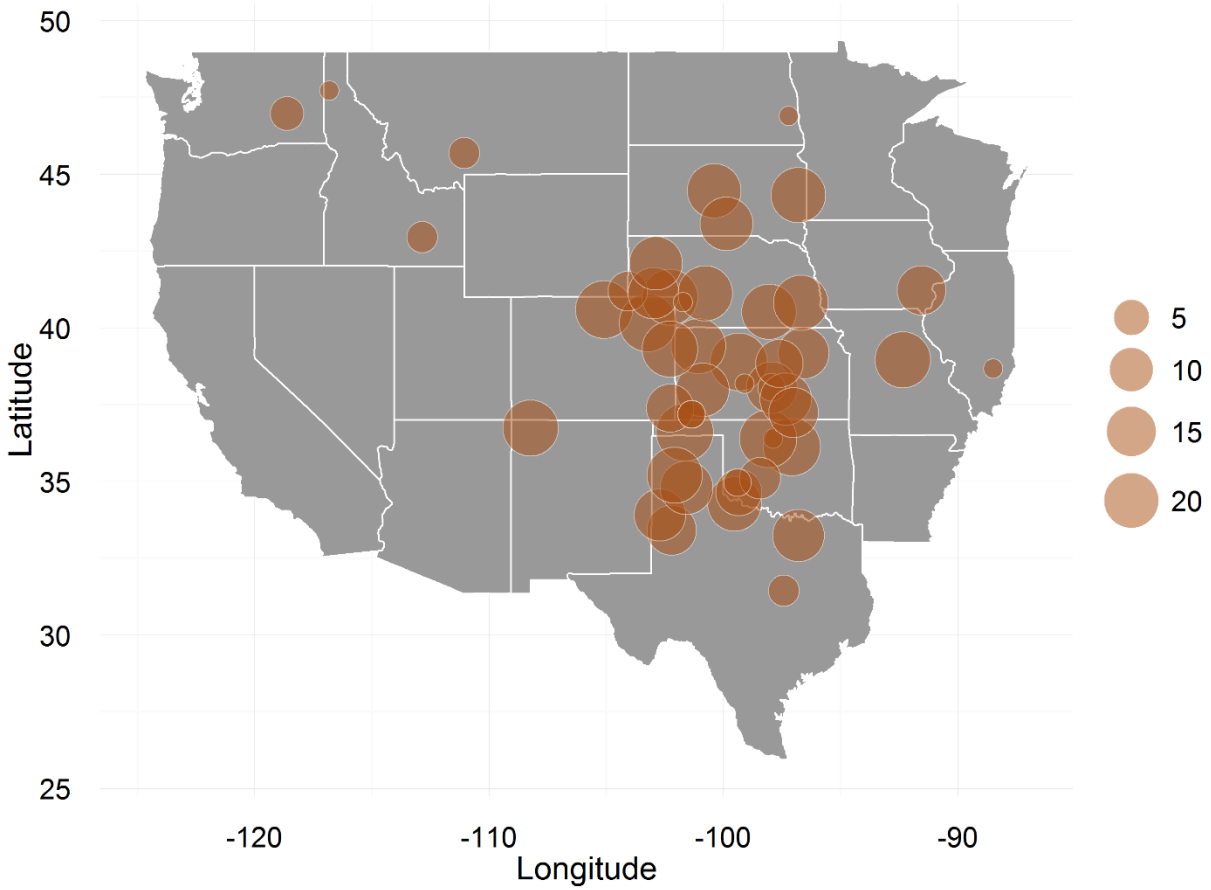

Figure 3-1. A map of SRPN locations from 1992-2015. The size of each circle indicates how many years the location was included in the nursery, with a minimum of 1 and a maximum of 23.

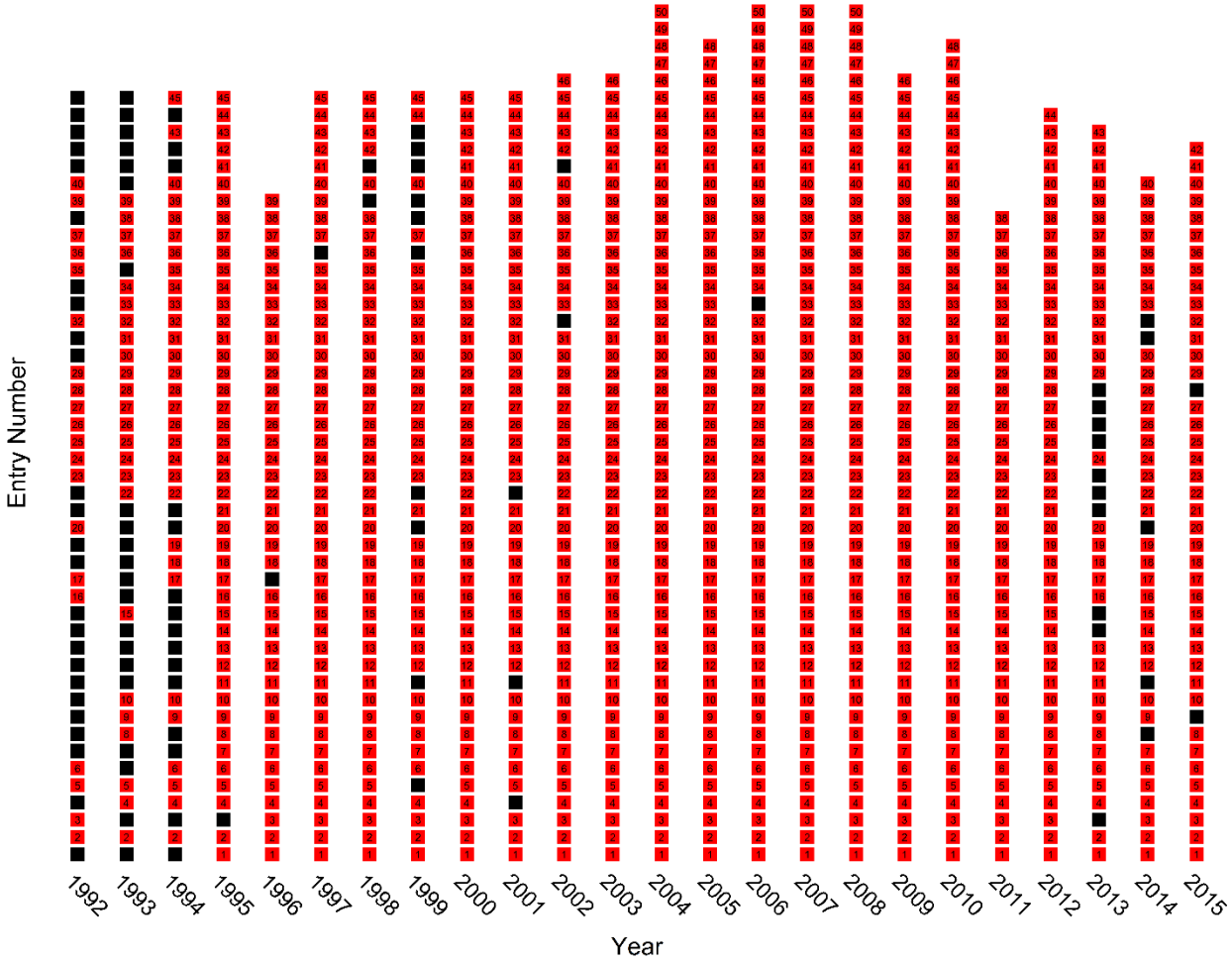

Figure 3-2. A visual representation of entries in the SRPN from 1992-2015. Each box represents an entry that was submitted to the SRPN. Red boxes indicate entries for which genotypic data was utilized in this experiment.

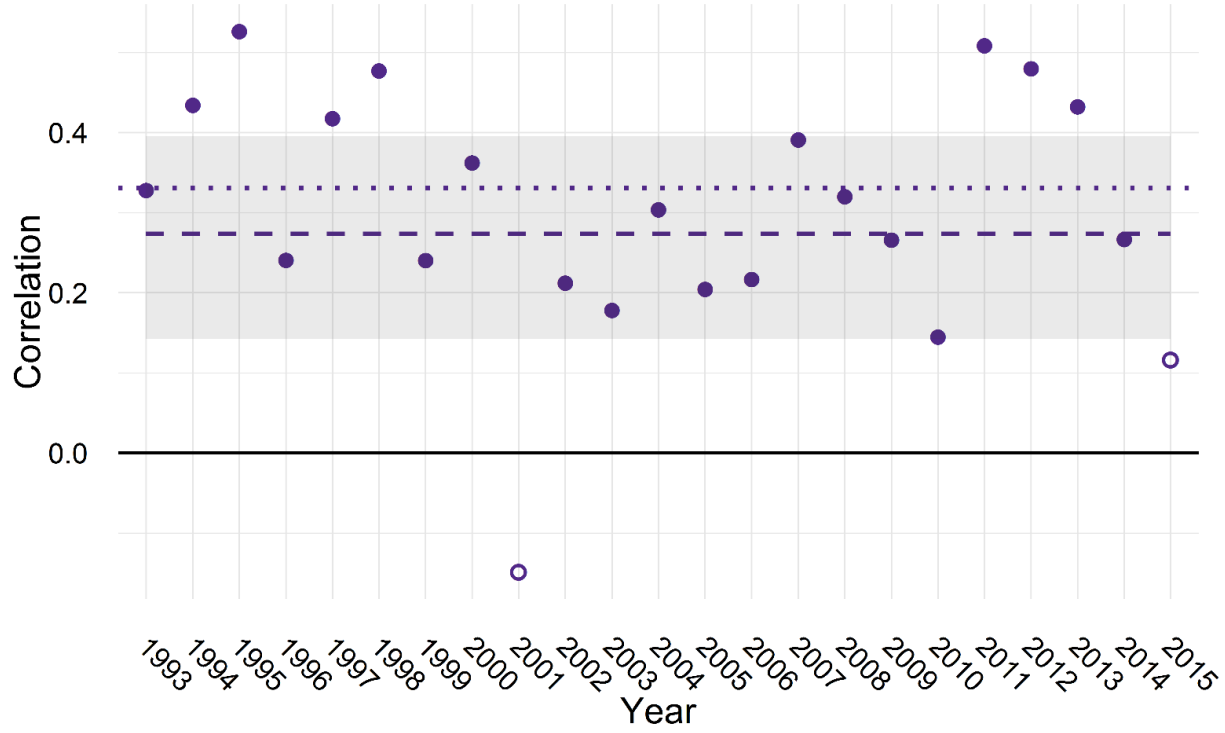

Figure 3-3. The prediction accuracy when using all prior years to predict a given year. The dashed line indicates the calculated phenotypic correlation (r=0.27) of lines that were tested across multiple years. The shaded area indicates the 95% confidence interval of the phenotypic correlation. The dotted line indicates the average genomic prediction correlation (r=0.331). Filled circles indicate years that were included in the training population; open circles indicate years that were excluded from the training population.

AgriPro	79	0.359	0.331
ARS-Lincoln	22	0.199	0.339
Check	35	0.235	0.06
Colorado State University	90	0.375	0.38
Kansas State University	85	0.279	0.325
Kansas State University - Hays	48	0.603	0.537
Limagrain Cereal Seeds	19	0.219	0.275
Monsanto	62	0.217	0.244
Oklahoma State University	123	0.283	0.354
Texas A&M University	17	0.777	0.664
Texas A&M University - Amarillo	59	0.083	0.03
Texas A&M University - Dallas	30	0.046	0.03
Texas A&M University - Vernon	66	0.416	0.43
Trio Research Inc.	67	0.282	0.462
University of Nebraska	96	0.084	0.138

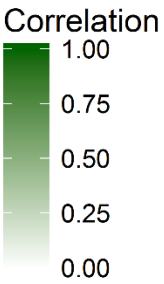

Figure 3-4. The prediction accuracies for individual breeding programs. Each row contains the name of the breeding program, the number of lines used in the analysis, the correlation when using a training program comprised of all lines (Left), and the correlation when using only lines originating from the same breeding program (Right).

Table 3-1. Number of lines tested and Nei's genetic distance for breeding programs submitting more than ten entries to the Southern Regional Performance Nursery from 1992 – 2015. The diversity estimate across the entire collection was 0.264.

Diversity	Program	Lines
0.197	Kansas State University - Hays	41
0.198	Trio Research Inc.	64
0.200	Texas A&M University - Dallas	18
0.212	ARS-Lincoln	19
0.217	Texas A&M University	15
0.220	Colorado State University	83
0.220	University of Nebraska	65
0.222	Texas A&M University - Amarillo	43
0.226	Texas A&M University - Vernon	54
0.226	Limagrain Cereal Seeds	20
0.242	Kansas State University	71
0.244	AgriPro	71
0.245	Oklahoma State University	99
0.250	Monsanto	65

Chapter 4 - Spiked GBS: A unified, open platform for single marker genotyping and whole-genome profiling

This chapter was adapted from the following peer-reviewed journal article:

Rife, T.W., S. Wu, R. Bowden, and J.A. Poland. 2015. Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. BMC Genomics 16(1): 1–7.

Abbreviations

MAS: marker-assisted selection; GS: genomic selection; KASP: Kompetitive Allele Specific PCR; TAS: Targeted amplicon sequencing; GBS: genotyping-by-sequencing

Abstract

Background

In plant breeding, there are two primary applications for DNA markers in selection: 1) selection of known genes using a single marker assay (marker assisted selection; MAS); and 2) whole-genome profiling and prediction (genomic selection; GS). Typically, marker platforms have addressed only one of these objectives.

Results

We have developed spiked genotyping-by-sequencing (sGBS), which combines targeted amplicon sequencing with reduced representation genotyping-by-sequencing. To minimize the cost of targeted assays, we utilize a small percent of available sequencing capacity available in runs of GBS libraries to "spike" amplified targets of *a priori* alleles tagged with a different set of unique barcodes. This open platform allows multiple, single-target loci to be assayed while simultaneously generating a whole-genome profile. This dual-genotyping approach allows different sets of samples to be evaluated. Here, we report the application of sGBS on a winter

wheat panel that was screened for converted KASP markers and newly-designed markers targeting known polymorphisms in the leaf rust resistance gene *Lr34*.

Conclusions

The flexibility and low-cost of sGBS will enable a range of applications across genetics research. Specifically, in breeding applications, the sGBS approach will allow breeders to obtain a whole-genome profile of important individuals while simultaneously targeting specific genes for a range of selection strategies across the breeding program.

Background

Progress in plant breeding focuses on the rapid development of new cultivars with improved attributes. Molecular markers allow breeders to characterize specific lines without the need for laborious and time-consuming phenotyping. Marker-assisted selection (MAS) is used in plant breeding to identify the allele present at a specific locus, allowing the breeder to select based on genotype (Collard et al., 2005). MAS has been used for plant breeding in many crops to identify specific individuals with known genes of interest (Buerstmayr et al., 2009; Suh et al., 2011; Zhao et al., 2012), primarily to target large-effect, single targets (Xu and Crouch, 2008; Collard and Mackill, 2008). Since each locus is generally genotyped independently, breeders tend to consider *per data point* costs when utilizing MAS within breeding programs.

Contemporary marker technologies for assaying single targets that are often used with MAS include KASP, targeted amplicon sequencing, and SNP arrays. KASP (Kompetitive Allele Specific PCR) is a uniplex, fluorescence-based single nucleotide genotyping technology that utilizes allele-specific oligo extension (Semagn et al., 2013). KASP markers have been used for breeding, QTL mapping, and are the main genotyping platform for the Generation Challenge Program at CIMMYT (Semagn et al., 2013). The arrival of inexpensive sequencing has led to the

development of economical sequence-based genotyping approaches. Targeted amplicon sequencing (TAS) amplifies known gene targets and attaches a barcode in a second PCR reaction for multiplexing (Bybee et al., 2011). Samples are pooled, sequenced, and analyzed by parsing the sample-specific barcode and then identifying sequence known or unknown variants (Durstewitz et al., 2010; Bybee et al., 2011). Using a targeted amplicon approach, Bybee et al. (Bybee et al., 2011) specifically looked at genes useful for phylogenetic analysis. TAS was further extended to a single PCR reaction that utilized linker sequences which allowed common target primers and a single set of barcoded primers to be utilized across distinct samples and loci (Clarke et al., 2014).

Complementary to assaying single loci for MAS, whole-genome profiling can be utilized for genomic section, QTL mapping, and diversity analysis (Jannink et al., 2010). Whole-genome profiling approaches focus on assaying large numbers of markers while reducing the *per sample* cost (Davey et al., 2011). Two common whole-genome profiling methods are SNP arrays and genotyping-by-sequencing (GBS). SNP arrays are comprised of a large number of known polymorphisms that allow an individual to be genotyped at all sites simultaneously which reduces the overall cost per data point (Ganal et al., 2012). SNP arrays have been used across a range of species to characterize diversity (Hyten et al., 2010b; Akhunov et al., 2010) and association mapping (Cockram et al., 2010). SNP arrays tend to be robust marker platforms but can have limitations, including the inability to target loci that were not included during the array development (i.e. ascertainment bias) and a relatively high per-sample cost.

GBS is a reduced representation whole-genome profiling strategy that leverages rapidly dropping sequencing cost and increasing output. Multiplexing samples with DNA barcodes greatly reduces the per sample cost (Elshire et al., 2011; Wetterstrand, 2014). GBS is one of

several reduced representation marker platforms to take advantage of second-generation sequencing platforms which produce enormous amounts of sequence (Davey et al., 2011; Poland and Rife, 2012). However, since many samples are sequenced together to minimize cost, the reduced sequencing coverage per sample often results in higher levels of missing data. Since sequencing is only targeted to regions flanking restriction sites, GBS is unable to directly ascertain specific loci, leading to considerable informatics challenges when used in MAS.

Spiked genotyping-by-sequencing (sGBS) takes advantage of abundant sequencing output by combining reduced representation GBS libraries with multiple, targeted amplicons. sGBS assesses known alleles via targeted amplicon sequencing and individual genotypes are determined by allele frequency counts. Multiple loci can be assayed concurrently since genotyping relies on the independent, raw sequence output. A similar approach to sGBS was developed by Wells et al. (Wells et al., 2013) that utilizes sequencing-based variant detection by barcoding amplicons. sGBS is distinguished that it is more economical since it uses only a small fraction of available sequencing capacity, the majority of which is simultaneously being used to generate independent, whole-genome profiles. By combining both approaches, breeders and geneticists are able to employ multi-faceted selection strategies and marker assays with a small increase in resource expenditure.

To evaluate this approach, we performed sGBS on a winter wheat panel that was screened for six converted KASP markers, four known polymorphisms in the leaf rust resistance gene *Lr34*, and one newly-designed marker targeting a known deletion in *Lr34*.

Methods

Plant Material

A panel of 153 diverse, advanced wheat lines (Table S1) was assembled and DNA was extracted from seedling leaf tissue using a BioSprint 96 DNA Plant Kit (Qiagen). DNA was quantified in plates using PicoGreen and concentrations were normalized to 20ng/µL.

Markers

Eleven single nucleotide markers were tested for the sGBS approach. Six of the markers were converted from a randomly chosen set of the KASP core markers: BS00023148, BS00083385, BS00150192, BS00067189, BS00088726, and BS00089969 (Wilkinson et al., 2012). Four of the markers were developed from previously designed *Lr34* KASP markers: Lr34exon11kasp, Lr34exon12kasp, Lr34intron4kasp, and Lr34exon22kasp (Lagudah et al., 2009). The 'Lr34exon11' marker from Lagudah et al. (Lagudah et al., 2009) was also adapted for sGBS, which targets a 3 bp insertion in exon 11, indicative of a non-functional allele (Lr34 minus). All primer and allele sequences are provided in Table S2. Two of the markers from the KASP core collection did not amplify (BS00067189 and BS00088726) and were not included in the subsequent analysis.

Primer Design

Primers were designed to amplify the full sequencing construct in a single PCR reaction (Figure 4-1). A set of 384 unique barcoded primers was developed for multiplexing and to differentiate spiked amplicons from GBS reads (Table S3). Each barcode primer contains the sequencer forward priming site, a unique 10-base barcode, and a M13 tail sequence (Figure 4-1). These were combined with locus-specific primers that also included the M13 tail sequence on the forward primer (Gholami et al., 2012). The locus-specific reverse primer includes both the

flanking sequence reverse primer and the sequencer-specific reverse priming site. Incorporating the M13 tail design on both the barcoded primer and allele-specific primer enables the utilization of the same set of barcode oligos for any target sequence, amortizing the cost of oligo synthesis across many samples. The alternative of making barcoded locus-specific primers for each target locus would be cost-prohibitive.

KASP markers were converted to primers for sGBS by removing the selective base on the end of each forward primer, effectively creating a single, common forward primer for each locus rather than the two allele specific primers used for KASP genotyping. Integrating the respective M13 and reverse Ion Torrent sequences on the primer pair made the KASP primer sequences compatible with sGBS.

Locus-Specific Amplification

In a 96 well plate, 150ng of DNA was combined with 3 pmol of M13 barcode primer (4µL at 0.75 µM). A master mix consisting of buffer (1X final), 0.75 µL $MgCl_2$ at 50 mM (2.5 mM final concentration), 1.2 µL dNTP mix at 2.5 mM for each nucleotide (200 µM final concentration for each), 0.3 pmol forward-tailed primer (0.03 µL at 10 µM: 20nM final concentration), 3 pmol reverse primer (0.3µL at 10 µM: 200nM final concentration), 0.33 U Taq polymerase, and 3.62µL H_2O were combined with the DNA for a total volume of 15µL for each reaction. Plates were PCR-amplified for 36 cycles consisting of 95C (1 min), 57C (20s), and 72C (40s). All samples in the plates were pooled and added to the quantified GBS libraries.

Library Construction and Sequencing

Two GBS libraries were prepared for Ion Torrent™ (Life Technologies, Carlsbad, CA) sequencing following the protocol from Mascher et al. (Mascher et al., 2013). Libraries were size-selected on a 2% agarose gel between 200 and 250bp, quantified using Quant-iTTM

PicoGreen® (Molecular Probes / Invitrogen Eugene, OR 97402), and normalized to 11nM. After pooling, the amplicon libraries were quantified using PicoGreen and normalized to 1.1nM. Five µL of the pooled amplicons were added to 50 µL of each GBS library for a final concentration of 1% (Figure 4-2). The libraries were prepared using the Ion PI™ Template OT2 200 Kit (v2 and v3) and then sequenced on an Ion Proton™ System using the Ion PI™ Chip Kit v1. The full protocol for library preparation is provided in Appendix C.

Data Processing

A TASSEL pipeline designed for Illumina sequence data was modified to identify SNPs from the GBS tags (Mascher et al., 2013; Glaubitz et al., 2014). Specifically, TASSEL was modified to process Ion Torrent sequencing sites and work with variable length sequence reads. SNP genotypes were called according to the approach of Poland et al. (Poland et al., 2012b) using a population-based filter. A TASSEL-based custom pipeline was written to determine the allele counts at each amplified locus by identifying the presence of both the M13 sequence and the target SNP alleles. Reads with the M13 tail sequence were parsed by barcode and the number of reads at each allele for a given locus was counted by exact matching to one of the target sequences.

Genotype calling for locus-specific amplicons

Lines with less than 10x read coverage were not included when clustering and calling genotypes. Genotypes were called using k-means clustering and DBSCAN clustering, both performed in R (Ester et al., 1996; Hennig, 2014; R Core Team, 2014). For k-means, the relative proportion of reads for each allele were plotted to determine the appropriate number of clusters to use for this input parameter. DBSCAN relies on reachability distance to determine the appropriate number of clusters (Ester et al., 1996; Hennig, 2014). Varying reachability distances

were empirically tested to ascertain an appropriate value. Observationally, a reachability distance of 0.1 ideally grouped all but one locus. For BS00150192, the optimal reachability distance was 0.06.

Results and Discussion

To test the approach of spiked GBS, we assayed a panel of diverse wheat lines using GBS to create a whole-genome profile and sGBS to target 11 known polymorphic sites. DNA was extracted and normalized and GBS libraries were constructed for the Ion Proton sequencing platform. The two sequenced GBS libraries contained 73M and 81M reads with a respective mean read length of 145bp and 183bp. Consistent with previous experience with unspiked GBS libraries, 83.6% and 81.3% of reads contained a good GBS barcode and a barcode plus enzyme cut site, respectively. Internal alignment-based discovery resulted in the identification of 13,617 SNPs with less than 20% missing data, also consistent with previous unspiked GBS libraries (Poland et al., 2012a; Mascher et al., 2013).

As a proportion of total sequencing output, the spiked amplicons constituted 1.8% and 3.1% of each library as determined by a count of M13 sequences. Amplicon libraries were individually analyzed to avoid bias due to read number differences. For each locus, the allelic state of each line was determined by counting the number of reads containing both the sample-specific barcode and a given allele. Genotypes were called using k-means clustering in R and DBSCAN clustering using the fpc package in R (Ester et al., 1996; Hennig, 2014). Relative read frequency was used to group individuals into one of three classes: A, B, or Heterozygous. K-means requires a parameter specifying the number of expected clusters while DBSCAN requires the reachability distance (Ester et al., 1996). Both of these values require individual curation for loci to ensure two (A/H or A/B) or three (A/B/H) clusters are correctly called.

Generally, there were few differences in the results from either method. For single-copy loci, both methods performed equally well and homozygotes and heterozygotes were easily identifiable (Figure 4-3a). Loci with non-zero axis clusters were also easily identified with both methods. Clusters arising from multi-copy loci were often distinct enough to confidently postulate the genotype allelic state (Figure 4-3c). Overall, the level of concordance between the two clustering algorithms was high with 97.2% of the genotype calls the same between the two methods (Figure 4-3b,d). The majority of discordance was due to k-means requiring that all genotypes be classified whereas DBSCAN did not classify individuals outside of the main clusters. The DBSCAN algorithm is therefore likely of more use in polyploid species where a heterozygote may not be as readily identified (Figure 4-3d). Ignoring the individuals that DBSCAN did not classify, there was 100% agreement between the two methods.

Robust conversion of SNP markers between different platforms is important for future genotyping applications, but success can vary considerably (Ragoussis, 2006; Hyten et al., 2010a; Uitdewilligen et al., 2013). In this study, we observed a good level of conversion from the KASP markers. Two attempted primer sets did not result in amplifying the target sequence and further efforts to optimize conditions for these primer sets were not attempted. For markers that successfully amplified, the average call rate was 94.8%. Several markers from the KASP core set resulted in non-zero axis read count clusters, likely due to the existence of homologous copies of the target locus. The percentage of alleles called for each locus and average coverage are reported in Table 4-1.

Conclusions

With sGBS, we have developed a low-cost, flexible platform for whole-genome profiling and targeted, single-locus genotyping. The open architecture of primer design for the spiked

amplicons enables simple inclusion of new or different target loci. Utilizing a unique set of barcodes combined with locus-specific M13 tail primers enabled sequencing of amplified targets in parallel with GBS libraries. While GBS provides a very low-cost approach for whole-genome profiling, it relies on reproducibly sequencing between restriction sites and cannot target *a priori* selected loci. Targeted amplicons fill this gap by allowing specific loci to be characterized. However, with the enormous sequencing output from current sequencing platforms, generating a sufficient number of amplicons across an appropriate number of samples to avoid unreasonable sequencing depth and cost is prohibitive. To minimize cost, we utilize a small fraction of the sequencing run (1-3%) while generating more than sufficient coverage across all target loci. Any reasonable number of amplicons could likely be combined with a GBS run. As with any sequencing approach, increasing the number of samples (or targets) decreases coverage. As sequencing output continues to increase, further 'excess' capacity can be leveraged in this way. However, as noted, targeted amplicon numbers beyond 10-20 are likely to be impractical relative to a fully designed array or whole-genome characterization (i.e. GBS).

Routine implementation of genotyping approaches in large genetic and breeding applications requires simple and robust laboratory pipelines. In concert with GBS library development, sGBS target amplification is a streamlined procedure affording routine, high-throughput implementation. The amplicon libraries are generated through a single PCR reaction, collectively normalized, and pooled with a GBS library. Though not attempted here, multiplex PCR reactions for the locus-specific amplification would further simplify the overall protocol.

sGBS was designed for MAS and GS in breeding but is also broadly applicable for a large number of other molecular genetics purposes. Many approaches ranging from diversity studies (Lu et al., 2013) to genetic and association mapping (Liu et al., 2014) and genomic

selection (Poland et al., 2012b) have successfully applied GBS, but the number of genetic markers generated by GBS often exceeds what is needed for genetic studies, such as fine mapping or TILLING. Fine mapping for map-based cloning generally requires screening a very large population with two flanking markers for the gene of interest. While GBS is not a suitable marker platform for fine mapping, utilizing the spiked portion of sGBS for these studies would be ideal. Likewise, the targeted amplicons of sGBS could also be used to screen for novel mutations in TILLING or ECO-TILLING populations. Though *a priori* SNPs were targeted in the present study, the direct sequencing of targets also enables *de novo* discovery of novel mutations as in a TILLING study.

For plant breeding, sGBS will enable breeders to genotype large collections of germplasm for specific markers by taking advantage of the massive data output of current sequencing platforms. Large numbers of markers are required for genomic selection, but plant breeders are also interested in characterizing important disease or physiological loci in breeding populations. sGBS provides a low-cost, scalable approach for both requirements and will serve as an important tool as plant breeding continues its use of molecular markers.

Since sGBS amplicons are independent of GBS libraries, breeders can generate a whole-genome profile for advanced breeding material while also applying marker-assisted selection to earlier generations. Importantly, the only realized cost for target genotyping using sGBS is a single PCR reaction. The ability to quickly identify lines containing specific alleles will enhance the capacity and speed of superior cultivar generation in breeding programs.

Plant breeding is inherently an exercise in producing and analyzing large amounts of data to discover improved rare and novel variants. Future advancements in plant breeding will fundamentally rely on new technologies being implemented that allow breeders to progress

through this process with the most efficient utilization of resources and least disruption to current workflow. Plant breeding programs have historically depended on single-marker germplasm characterization and are beginning to take advantage of whole-genome profiles for genomic selection. sGBS combines both approaches, eliminating the current necessity of two distinct platforms while leveraging continual advancements in sequencing technology. This efficient strategy will allow breeders to increase the amount of germplasm and number of loci that are assayed with few changes to workflow and limited expenditure of resources. Developments like sGBS that will enable genomics-assisted breeding are crucial to ensuring progress in developing improved plant varieties in the effort to eliminate hunger and poverty across the world.

Supporting Data

1. Sequence files archived at NCBI SRA under accession number SRP052305.
2. Supplemental Table S1. Wheat varieties used in this analysis.
3. Supplemental Table S2. Loci, target alleles, and primer sequences used for sGBS.
4. Supplemental Table S3. Barcode sequences and forward oligo sequences.
5. Supplemental File S4. Full protocol for spiked genotyping-by-sequencing.

Acknowledgements

We would like to thank the USDA Central Small Grain Genotyping Lab in Manhattan, KS for sequencing. The USDA-NIFA funded Triticeae Coordinated Agriculture Project (T-CAP) (2011-68002-30029) provided support for TR. This work was completed under the auspices of WGRC I/UCRC partially funded by NSF grant contract (IIP-1338897) and the USAID Feed the Future Innovation Lab for Applied Wheat Genomics (Cooperative Agreement No. AID-OAA-A-13-00051). Funding for this research was provided by the Bill & Melinda Gates Foundation through a grant to Cornell University for “Genomic Selection: The next frontier for rapid gains

in maize and wheat improvement" and the United States Department of Agriculture-Agricultural Research Service (Appropriation #5430-21000-006-00D). Additional support was provided through the Kansas Wheat Alliance and the Kansas Wheat Commission. Mention of trade names does not constitute endorsement by the U.S. Department of Agriculture.

References

Akhunov, E.D., A.R. Akhunova, O.D. Anderson, J. a Anderson, N. Blake, M.T. Clegg, D. Coleman-Derr, E.J. Conley, C.C. Crossman, K.R. Deal, J. Dubcovsky, B.S. Gill, Y.Q. Gu, J. Hadam, H. Heo, N. Huo, G.R. Lazo, M.-C. Luo, Y.Q. Ma, D.E. Matthews, P.E. McGuire, P.L. Morrell, C.O. Qualset, J. Renfro, D. Tabanao, L.E. Talbert, C. Tian, D.M. Toleno, M.L. Warburton, F.M. You, W. Zhang, and J. Dvorak. 2010. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. BMC Genomics 11(1): 702.

Buerstmayr, H., T. Ban, and J.A. Anderson. 2009. QTL mapping and marker-assisted selection for Fusarium head blight resistance in wheat: a review. Plant Breed. 26(128): 1–26.

Bybee, S.M., H. Bracken-Grissom, B.D. Haynes, R.A. Hermansen, R.L. Byers, M.J. Clement, J.A. Udall, E.R. Wilcox, and K.A. Crandall. 2011. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol. Evol. 3: 1312–1323.

Clarke, L.J., P. Czechowski, J. Soubrier, M.I. Stevens, and A. Cooper. 2014. Modular tagging of amplicons using a single PCR for high-throughput sequencing. Mol. Ecol. Resour. 14(1): 117–121.

Cockram, J., J. White, D.L. Zuluaga, D. Smith, J. Comadran, M. Macaulay, Z. Luo, M.J. Kearsey, P. Werner, D. Harrap, C. Tapsell, H. Liu, P.E. Hedley, N. Stein, D. Schulte, B. Steuernagel, D.F. Marshall, W.T.B. Thomas, L. Ramsay, I. Mackay, D.J. Balding, R. Waugh, and D.M. O'Sullivan. 2010. Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc. Natl. Acad. Sci. U. S. A. 107(50): 21611–6.

Collard, B.C.Y., M.Z.Z. Jahufer, J.B. Brouwer, and E.C.K. Pang. 2005. An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142(1–2): 169–196.

Collard, B.C.Y., and D.J. Mackill. 2008. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 363(1491): 557–572.

Davey, J.W., P.A. Hohenlohe, P.D. Etter, J.Q. Boone, J.M. Catchen, and M.L. Blaxter. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing.

Nat. Rev. Genet. 12(7): 499–510.

Durstewitz, G., A. Polley, J. Plieske, H. Luerssen, E.M. Graner, R. Wieseke, and M.W. Ganal. 2010. SNP discovery by amplicon sequencing and multiplex SNP genotyping in the allopolyploid species Brassica napus. Genome 53(11): 948–956.

Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 6(5): e19379.

Ester, M., H. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. p. 226–231. *In* 2nd International Conference Knowledge Discovery in Databases and Data Mining.

Ganal, M.W., A. Polley, E.-M. Graner, J. Plieske, R. Wieseke, H. Luerssen, and G. Durstewitz. 2012. Large SNP arrays for genotyping in crop plants. J. Biosci. 37(5): 821–828.

Gholami, M., W.A. Bekele, J. Schondelmaier, and R.J. Snowdon. 2012. A tailed PCR procedure for cost-effective, two-order multiplex sequencing of candidate genes in polyploid plants. Plant Biotechnol. J. 10(6): 635–645.

Glaubitz, J.C., T.M. Casstevens, F. Lu, J. Harriman, R.J. Elshire, Q. Sun, and E.S. Buckler. 2014. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. PLoS One 9(2): e90346.

Hennig, C. 2014. fpc: Flexible procedures for clustering.

Hyten, D.L., S.B. Cannon, Q. Song, N. Weeks, E.W. Fickus, R.C. Shoemaker, J.E. Specht, A.D. Farmer, G.D. May, and P.B. Cregan. 2010a. High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11: 38.

Hyten, D.L., I.-Y. Choi, Q. Song, J.E. Specht, T.E. Carter, R.C. Shoemaker, E.-Y. Hwang, L.K. Matukumalli, and P.B. Cregan. 2010b. A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. Crop Sci. 50(3): 960–968.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9(2): 166–177.

Lagudah, E.S., S.G. Krattinger, S. Herrera-Foessel, R.P. Singh, J. Huerta-Espino, W. Spielmeyer, G. Brown-Guedira, L.L. Selter, and B. Keller. 2009. Gene-specific markers for the wheat gene Lr34/Yr18/Pm38 which confers resistance to multiple fungal pathogens. Theor. Appl. Genet. 119(5): 889–898.

Liu, H., M. Bayer, A. Druka, J.R. Russell, C.A. Hackett, J. Poland, L. Ramsay, P.E. Hedley, and R. Waugh. 2014. An evaluation of genotyping by sequencing (GBS) to map the Breviaristatum-e (ari-e) locus in cultivated barley. BMC Genomics 15: 104.

Lu, F., A.E. Lipka, J. Glaubitz, R. Elshire, J.H. Cherney, M.D. Casler, E.S. Buckler, and D.E. Costich. 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. PLoS Genet. 9(1): e1003215.

Mascher, M., S. Wu, P.S. Amand, N. Stein, and J. Poland. 2013. Application of Genotyping-by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and Reference-Based Marker Ordering in Barley. PLoS One 8(10): e76925.

Poland, J.A., P.J. Brown, M.E. Sorrells, and J.-L. Jannink. 2012a. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7(2): e32253.

Poland, J.A., J. Endelman, J. Dawson, J. Rutkoski, S. Wu, Y. Manes, S. Dreisigacker, J. Crossa, H. Sánchez-Villeda, M. Sorrells, and J.-L. Jannink. 2012b. Genomic Selection in Wheat Breeding using Genotyping-by-Sequencing. Plant Genome 5(3): 103–113.

Poland, J.A., and T.W. Rife. 2012. Genotyping-by-Sequencing for Plant Breeding and Genetics. Plant Genome J. 5(3): 92–102.

R Core Team. 2014. R: A language and environment for statistical computing.

Ragoussis, J. 2006. Genotyping technologies for all. Drug Discov. Today Technol. 3(2): 115–122.

Semagn, K., R. Babu, S. Hearne, and M. Olsen. 2013. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol. Breed. 33(1): 1–14.

Suh, J.-P., S.-J. Yang, J.-U. Jeung, A. Pamplona, J.-J. Kim, J.-H. Lee, H.-C. Hong, C.-I. Yang, Y.-G. Kim, and K.K. Jena. 2011. Development of elite breeding lines conferring Bph18 gene-derived resistance to brown planthopper (BPH) by marker-assisted selection and genome-wide background analysis in japonica rice (Oryza sativa L.). F. Crop. Res. 120(2): 215–222.

Uitdewilligen, J.G.A.M.L., A.-M.A. Wolters, B.B. D'hoop, T.J.A. Borm, R.G.F. Visser, and H.J. van Eck. 2013. A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. PLoS One 8(5): e62355.

Wells, R., M. Trick, F. Fraser, E. Soumpourou, L. Clissold, C. Morgan, J. Pauquet, and I. Bancroft. 2013. Sequencing-based variant detection in the polyploid crop oilseed rape. BMC Plant Biol. 13(1): 111.

Wetterstrand, K.A. 2014. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Available at www.genome.gov/sequencingcosts (verified 21 April 2014).

Wilkinson, P.A., M.O. Winfield, G.L.A. Barker, A.M. Allen, A. Burridge, J.A. Coghill, and K.J. Edwards. 2012. CerealsDB 2.0: an integrated resource for plant breeders and scientists. BMC Bioinformatics 13: 219.

Xu, Y., and J.H. Crouch. 2008. Marker-Assisted Selection in Plant Breeding: From Publications to Practice. Crop Sci. 48(2): 391–407.

Zhao, X., G. Tan, Y. Xing, L. Wei, Q. Chao, W. Zuo, T. Lübberstedt, and M. Xu. 2012. Marker-assisted introgression of qHSR1 to improve maize resistance to head smut. Mol. Breed. 30(2): 1077–1088.

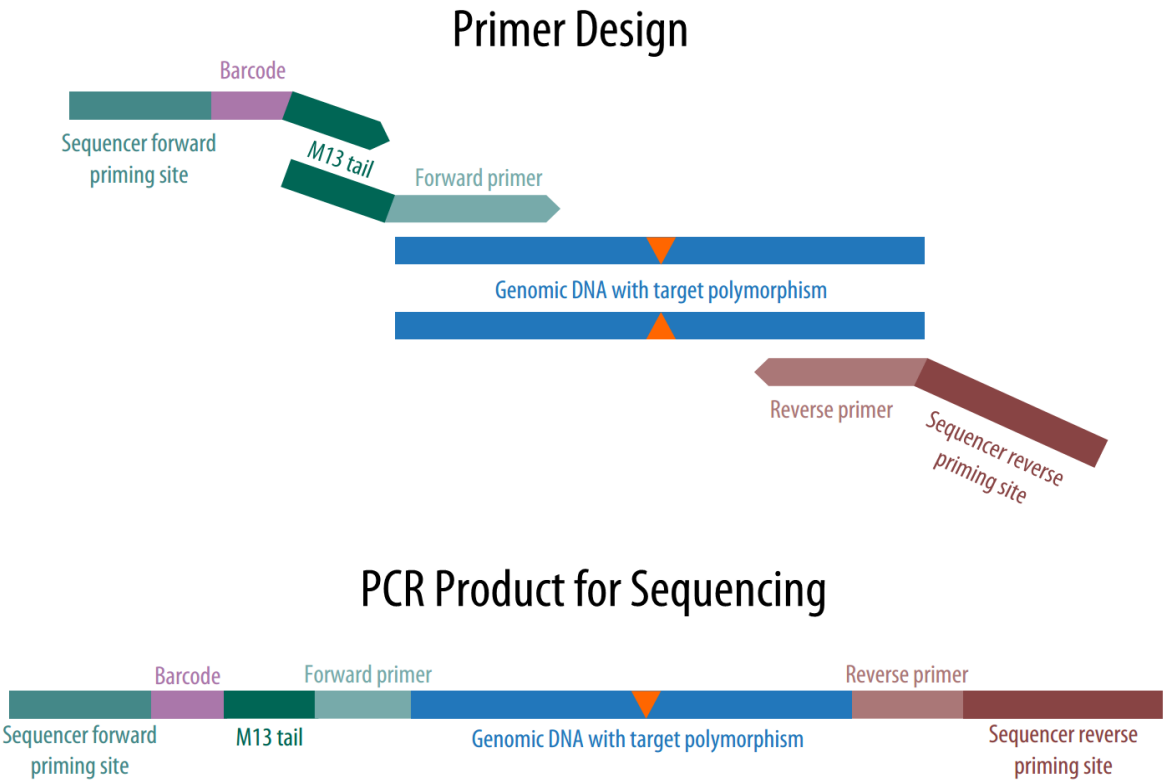

Figure 4-1. Primer and amplicon construction. The first round of PCR uses a forward primer containing the M13 sequence to amplify the target region. The second round of PCR extends from the M13 tail and incorporates a unique barcode, leading to a final product containing the sequencer primers, barcode, M13 sequence, and polymorphic target.

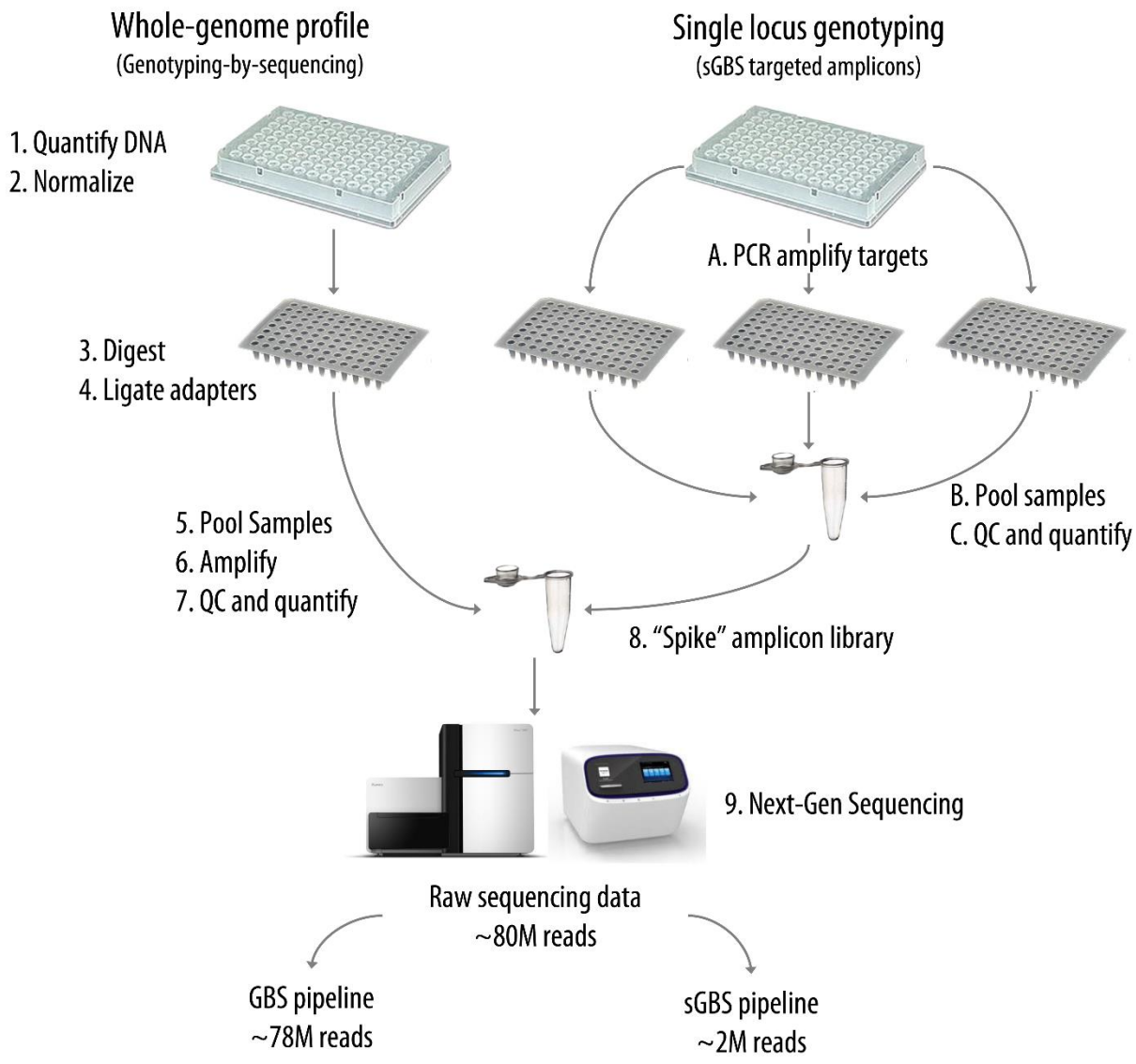

Figure 4-2. Library construction flow chart. GBS libraries are created following standard protocols. Each spiked library amplifies a single target locus. Spiked libraries are pooled, combined with GBS libraries, and sequenced. Sequence data for the amplicon library is parsed using the M13 and unique barcode sequence.

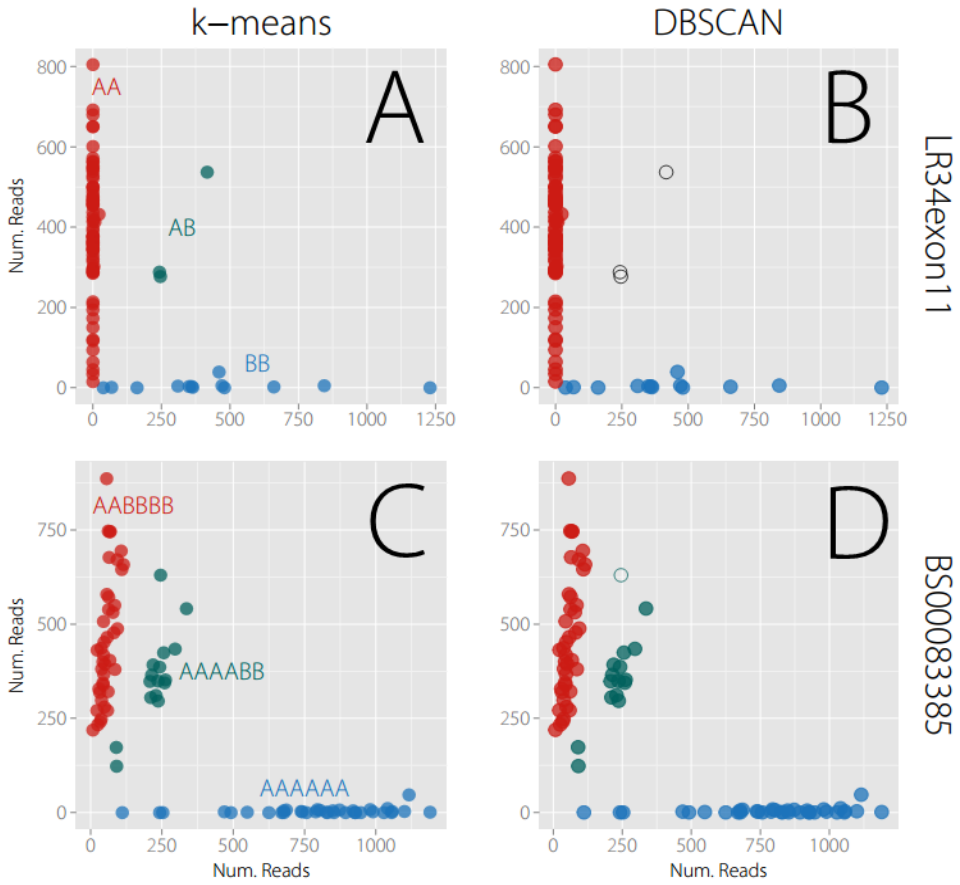

Figure 4-3. k-means clustering and DBSCAN clustering for Lr34exon11 and BS00083385. k-means clustering and DBSCAN clustering were used to cluster genotypes for each individual on relative read frequency of the two SNP alleles. Genotypes called within the same group are denoted by color. Unfilled symbols indicate samples that were not classified by the algorithms. (A) k-means and (B) DBSCAN clustering of LR34exon11. LR34exon11 locus is a single-copy locus and the two genotypes are easily distinguished by either clustering algorithm. Heterozygotes are characterized by an equal proportion of both alleles. (C) k-means and (D) DBSCAN clustering of BS00083385. This primer set presumably amplifies multiple loci in the polyploid wheat genome that can still be distinguished based on relative read frequency. The three genotypic classes for individual lines are likely AAAAAA, AABBBB, and AAAABB. The BBBBBB group does not appear to be present as a null A genotype should fall on the vertical axis. (Zero reads counts of allele A.) DBSCAN did not classify the unfilled individual, which is potentially a heterozygous genotype at one of the loci (AAABBB).

Table 4-1. Marker name, total call rate, and average read depth.

Marker	Call Rate	Avg. Depth
LR34exon11	94.5%	336
Lr34intron4kasp	96.4%	114
Lr34exon12kasp	99.3%	923
LR34exon11kasp	98.7%	1573
Lr34exon22kasp	99.2%	117
BS00150192	92.8%	863
BS00089969	92.7%	564
BS00023148	98.2%	1577
BS00083385	81.0%	1118

Chapter 5 - An Open-Source Application for Field Data Collection on Android

This chapter has been published as the following peer-reviewed journal article:

Rife, T.W., and J.A. Poland. 2014. Field Book: An Open-Source Application for Field Data Collection on Android. Crop Sci. 54(4): 1624–1627.

Abstract

Plant breeding and genetics research is an inherently data-driven enterprise. Typical experiments and breeding nurseries can contain thousands of unique entries and programs will often evaluate tens of thousands of plots each year. To function efficiently on this scale, electronic data management becomes essential. Many research programs, however, continue to operate by scribing and transcribing massive amounts of data on paper field books. While effective, this form of data management places heavy burdens on human resources, decreases data integrity, and greatly limits future utilization of data and the ability to expand the breeding program. To help address these constraints, we have developed an open-source application for electronic data capture that runs on consumer-grade Android tablets. By focusing on a simple, stand-alone application with an intuitive and customized interface, we attempt to decrease both the technological and cost barriers that hinder adoption of electronic data management in breeding programs. The simplicity of Field Book allows adoption of the technology without a steep learning curve. With low-cost, accessible solutions, the vision of one handheld per breeder can become a reality for breeding programs around the world. Transformational capacity in electronic data collection and management will be essential to realize a contemporary green revolution.

Introduction

Accurate data collection is a fundamental requirement for plant science research and plant breeding where large populations are required for dissecting quantitative traits and selecting improved varieties (Falconer and Mackay, 1996). In accordance with quantitative genetic theory, it has been demonstrated that power for QTL and association mapping is a function of population size (Vales et al., 2005; Yu et al., 2008; Myles et al., 2009; Buckler et al., 2009). Further, genomic selection and applications of marker-assisted selection are no exception to the rule of larger populations. Larger population sizes lead to larger gain and better probability to identify superior candidate varieties (Jannink et al., 2010). The evaluation of large populations is often limited from a functional perspective by the ability to evaluate a large amount of genetic material.

While the generation of genetic data has undergone a high-throughput revolution, phenotypic evaluation of genetic populations and experimental lines remains time consuming and expensive. Tools that can be applied to increase the speed and efficiency of phenotypic evaluation will help generate high-value data from field trials. A typical field season requires considerable preparation since field data must be organized specifically for data collection. Once collected, thousands of data points must be transcribed, often by a dedicated employee, creating an enormous bottleneck in the workflow of the project, introducing transcription errors that can subsequently affect analysis, and requiring considerable investment of human resources (Easton et al., 2000).

Current electronic systems for data collection in field trials are often associated with proprietary, expensive hardware and software. This can prevent adoption and lock researchers into a single platform. Other platforms rely on the user to manually assemble specific hardware

(Berke and Baenziger, 1992). Adoption of new tools requires a significant investment, decreasing the likelihood that researchers will assess and investigate different platforms. Since proprietary hardware is often based on older technology, these platforms lag behind what is currently available to consumers. This leaves researchers with less flexible and less functional hardware.

Form and Function

We have designed and programmed Field Book, an open-source application that runs on Android. This application addresses many problems inherent to other field data collection software and paper field books. Field Book, including all source code, is freely available and developers can further customize the application to meet specific data collection requirements. The application and all associated documentation is available at the Poland Lab website (http://www.wheatgenetics.org/field-book) and the Google Play Store (http://play.google.com/store/apps/details?id=com.fieldbook.tracker&hl=en). Field Book runs on consumer-grade hardware (~$200) as well as more expensive, rugged tablets (~$1200). With inexpensive hardware, it becomes feasible to purchase a device for each person collecting data in the field.

Field Book was designed to display data at an individual entry level with the capacity to navigate independently between traits and entries. The interface is designed to facilitate easy and rapid data entry for one or multiple traits on each plot. “InfoBars” at the top of the screen can display additional imported data, allowing the user to have much more information available in the field than paper field books. Users can easily display entry names, pedigrees, seed sources, entry codes, or any additional records of interest for each field record (Figure 5-1).

Field Book employs multiple data input formats depending on the nature of the trait data being collected (Figure 5-2). Trait formats include numeric, categorical, Boolean, percent, date, text, and audio. At any time, the user can define new traits to be collected. Traits can be created, reordered, and removed from within Field Book. During data collection, the traits of interest can be selected and all others hidden so that the main screen will only display the traits needed at that specific time. During data entry, users can scroll at a plot level or a trait level. In this way, users can sequentially move through plots on a single trait entry, or move through multiple traits on a given plot before advancing to the next plot.

Data can be exported in either a database or table format. Database format exports each observation independently and can be immediately uploaded to and stored in a central database. This format includes all metadata for a given phenotypic measurement, such as the name of the person recording the data and the timestamp. Table export uses the traditional spreadsheet format with a list of entries in rows and columns corresponding to each trait. Database format is preferable for direct import to a relational database while the table format can quickly be imported into statistical software for analysis. Field Book allows both formats to be exported simultaneously.

There are a number of applications in managing plant breeding programs that can benefit from barcodes. Barcodes further increase speed in reading input data while reducing input errors. Field Book supports both wireless (via Bluetooth®) and wired (via USB On-The-Go) barcode scanners that can be used for data collection. In practice, traits have been successfully collected by scanning barcodes that correspond to a phenotypic value (e.g. plant height using a measuring stick with barcoded numbers) (Figure 5-3).

With an established framework for Field Book, additional features that appeal to a wider range of researchers are being added. Users are now able to visualize the state of data collection in the field with a field map. The map indicates which entries have and have not been collected for a specific trait and allows the user to perform a quick visual analysis on the data to identify outliers. Since the structure and flexibility of the software means that many devices will often be in the field at once, future updates will allow multiple devices to upload collected data to a central device, removing the need to manually combine or interact with multiple files from multiple devices.

Perspective

Field data collection is fundamental to plant breeding and genetics research. To strengthen field research in these programs, Field Book has been developed as an open-source tool that can be used to collect data on all kinds of experiments. The potential to increase the speed of collection and analysis will enable increases in the size of field experiments and, subsequently, the rate of genetic gain. The ability to keep data organized in digital form allows technicians and breeders to focus on other tasks, leading to further innovation and growth of plant breeding programs. The prospect of a contemporary green revolution is predicated on the development and production of improved, high-yielding varieties. To develop these improved varieties, a transformative implementation of electronic data capture and management in breeding programs will be critical. Field Book moves toward the vision of one handheld device per breeder, giving every breeder access to robust data collection and management that will facilitate the development of improved varieties to enable needed gains in agriculture productivity.

Acknowledgements

The development of Field Book was supported through The McKnight Foundation Collaborative Crop Research Program and USDA-ARS. Contracted support for programming Field Book was through Technology Projects (technologyprojects@gmail.com). Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

References

Berke, T., and P. Baenziger. 1992. Portable and desktop computer integrated field book and data collection system for agronomists. Agron. J. 84: 119–121.

Buckler, E.S., J.B. Holland, P.J. Bradbury, C.B. Acharya, P.J. Brown, C. Browne, E. Ersoz, S. Flint-Garcia, A. Garcia, J.C. Glaubitz, M.M. Goodman, C. Harjes, K. Guill, D.E. Kroon, S. Larsson, N.K. Lepak, H. Li, S.E. Mitchell, G. Pressoir, J. a Peiffer, M.O. Rosas, T.R. Rocheford, M.C. Romay, S. Romero, S. Salvo, H. Sanchez Villeda, H.S. da Silva, Q. Sun, F. Tian, N. Upadyayula, D. Ware, H. Yates, J. Yu, Z. Zhang, S. Kresovich, and M.D. McMullen. 2009. The genetic architecture of maize flowering time. Science (80-.). 325(5941): 714–718.

Easton, K.L., J.F. McComish, and R. Greenberg. 2000. Avoiding Common Pitfalls in Qualitative Data Collection and Transcription. Qual. Health Res. 10(5): 703–707.

Falconer, D.S., and T.F. Mackay. 1996. Introduction to Quantitative Genetics (4th edn). London, UK.

Jannink, J.-L., A.J. Lorenz, and H. Iwata. 2010. Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics 9(2): 166–177.

Myles, S., J. Peiffer, P.J. Brown, E.S. Ersoz, Z. Zhang, D.E. Costich, and E.S. Buckler. 2009. Association mapping: critical considerations shift from genotyping to experimental design. Plant Cell 21(8): 2194–2202.

Vales, M.I., C.C. Schön, F. Capettini, X.M. Chen, a E. Corey, D.E. Mather, C.C. Mundt, K.L. Richardson, J.S. Sandoval-Islas, H.F. Utz, and P.M. Hayes. 2005. Effect of population size on the estimation of QTL: a test using resistance to barley stripe rust. Theor. Appl. Genet. 111(7): 1260–1270.

Yu, J., J.B. Holland, M.D. McMullen, and E.S. Buckler. 2008. Genetic design and statistical power of nested association mapping in maize. Genetics 178(1): 539–551.

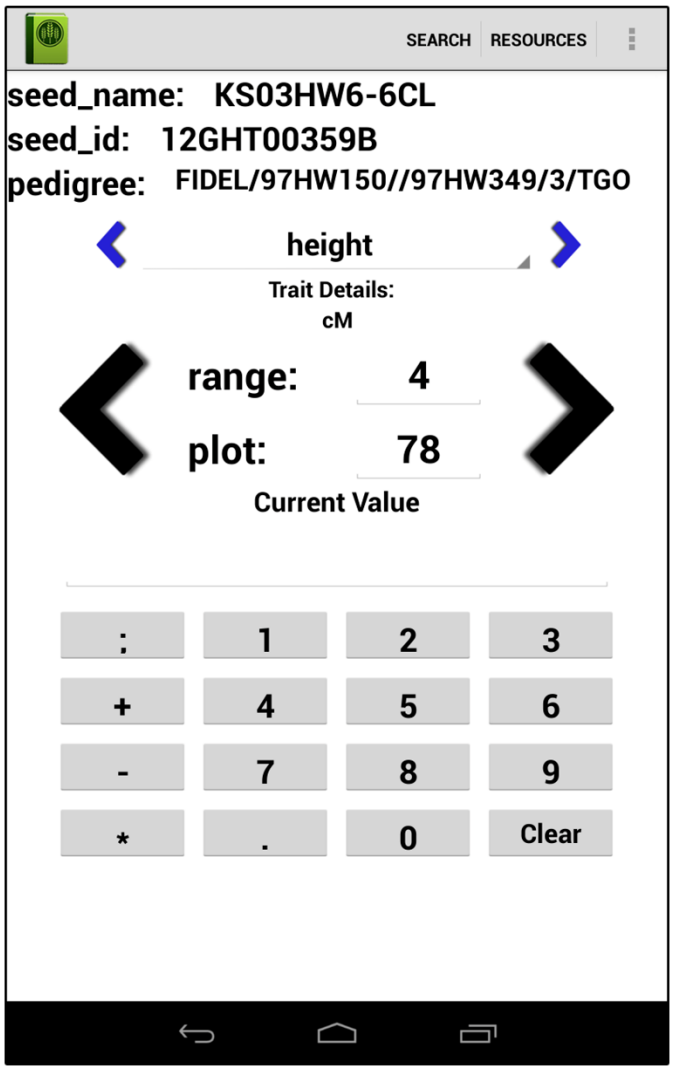

Figure 5-1. Main layout of Field Book application on an Android tablet. Three InfoBars (labeled seed_name, seed_id, and pedigree in this example) display additional information that was imported with the field file. The small, blue trait arrows allow the user to scroll through the different traits to be collected. The large, black arrows change the focus to the next or previous entry. The current value is displayed in the middle of the screen, and the bottom portion of the screen is reserved for data input. "Search" gives the user pseudo-query search capacity. "Resources" allows quick access to rating keys or field maps.

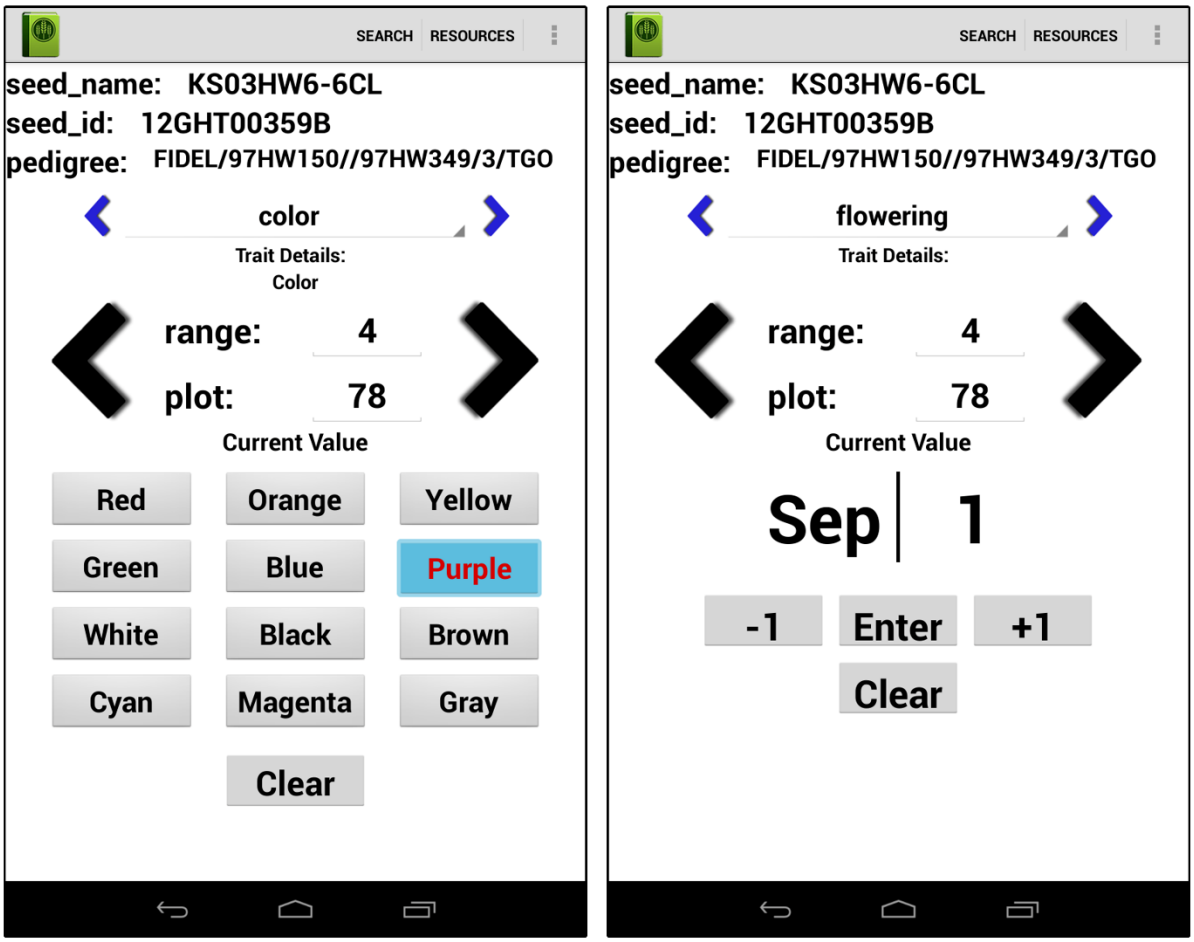

Figure 5-2. Examples of the custom input designed for categorical (left) and date (right) trait formats. For categorical traits, up to twelve categories can be defined for collection. Date format displays the current date with the option to increase or decrease that value by one day increments.

Figure 5-3. Collecting data by scanning the barcode corresponding to the height of the plot.

Appendix A - Copyright Permission

This appendix includes the copyright permissions and licenses required to republish the content in this dissertation.

ACSESS-ALLIANCE OF CROP, SOIL, AND ENVIRONMENTAL SCIENCE SOCIETIES LICENSE
TERMS AND CONDITIONS

Jul 25, 2016

This Agreement between Trevor Rife ("You") and ACSESS-Alliance of Crop, Soil, and Environmental Science Societies ("ACSESS-Alliance of Crop, Soil, and Environmental Science Societies") consists of your license details and the terms and conditions provided by ACSESS-Alliance of Crop, Soil, and Environmental Science Societies and Copyright Clearance Center.

License Number	3916060864707
License date	Jul 25, 2016
Licensed Content Publisher	ACSESS-Alliance of Crop, Soil, and Environmental Science Societies
Licensed Content Publication	Crop Science
Licensed Content Title	Field Book: An Open-Source Application for Field Data Collection on Android
Licensed Content Author	Trevor W. Rife, and Jesse A. Poland
Licensed Content Date	Jul 8, 2014
Licensed Content Volume Number	54
Licensed Content Issue Number	4
Type of Use	Thesis/Dissertation
Requestor type	Author of requested content
Format	Print, Electronic
Portion	chapter/article
Rights for	Main product
Creation of copies for the disabled	no
With minor editing privileges	no
For distribution to	Worldwide
In the following language(s)	Original language of publication
With incidental promotional use	no
The lifetime unit quantity of new product	0 to 499
The requesting person/organization is:	Trevor Rife
Order reference number	
Title of your thesis / dissertation	UTILIZING A HISTORICAL WHEAT COLLECTION TO DEVELOP NEW TOOLS FOR MODERN PLANT BREEDING
Expected completion date	Aug 2016
Estimated size (number of pages)	140

Requestor Location	Trevor Rife 4024 Throckmorton Hall Kansas State University MANHATTAN, KS 66502 United States Attn: Trevor Rife
Billing Type	Invoice
Billing Address	Trevor Rife 4024 Throckmorton Hall Kansas State University MANHATTAN, KS 66502 United States Attn: Trevor Rife
Total	0.00 USD

Terms and Conditions

Introduction

The Publisher for this copyrighted material is ACSESS. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your CCC account and that are available at any time at <http://myaccount.copyright.com>).

Limited License

Publisher hereby grants to you a non-exclusive license to use this material. Licenses are for one-time use only with a maximum distribution equal to the number that you identified in the licensing process; any form of republication must be completed within 60 days from the date hereof (although copies prepared before then may be distributed thereafter); and any electronic posting is limited to a period of 120 days.

Geographic Rights: Scope

Licenses may be exercised anywhere in the world.

Altering/Modifying Material: Not Permitted

You may not alter or modify the material in any manner, nor may you translate the material into another language.

Reservation of Rights

Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

License Contingent on Payment

While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

Copyright Notice: Disclaimer

Appendix B - Supplementary Materials Chapter 3

This appendix includes the supplementary figures and tables for the Chapter 3.

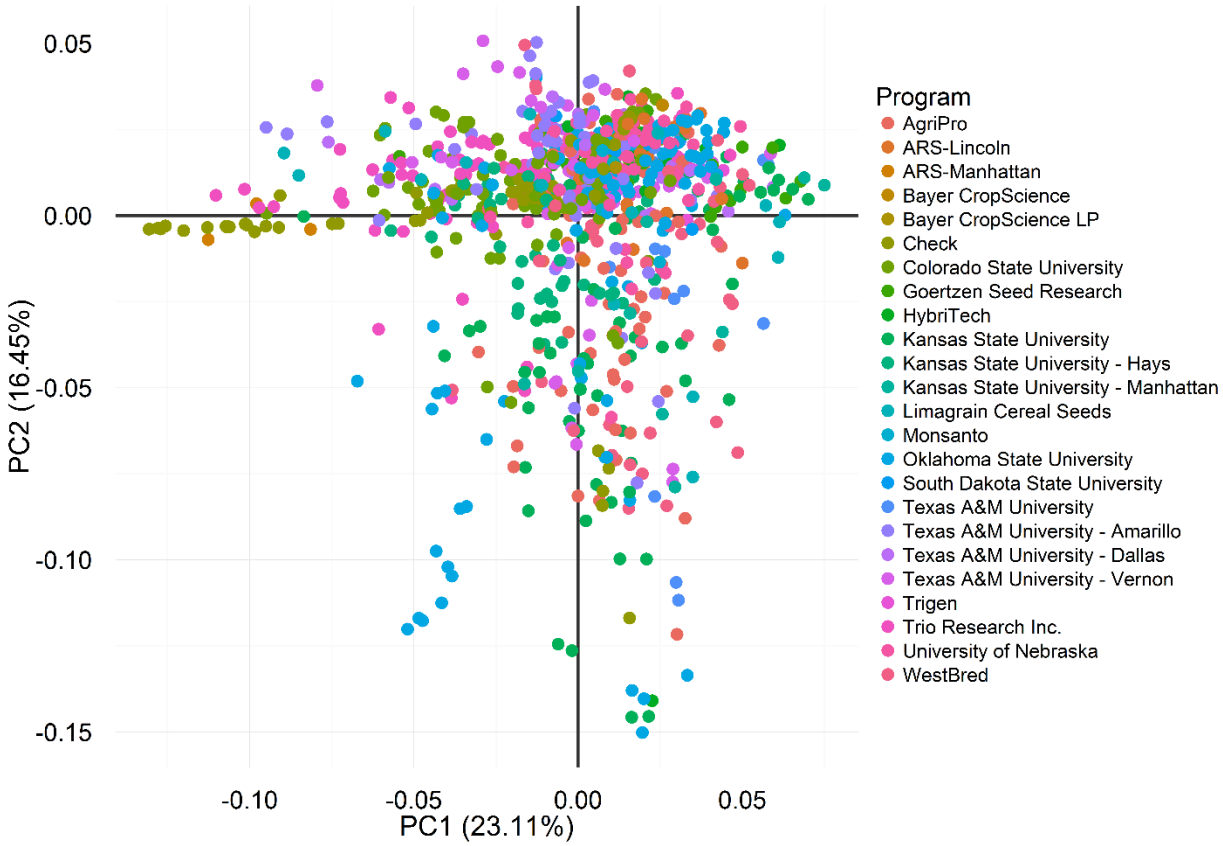

Figure 5-4. A plot of the first and second Eigen vectors derived from the A matrix using the eigen function in R (R Core Team, 2014).

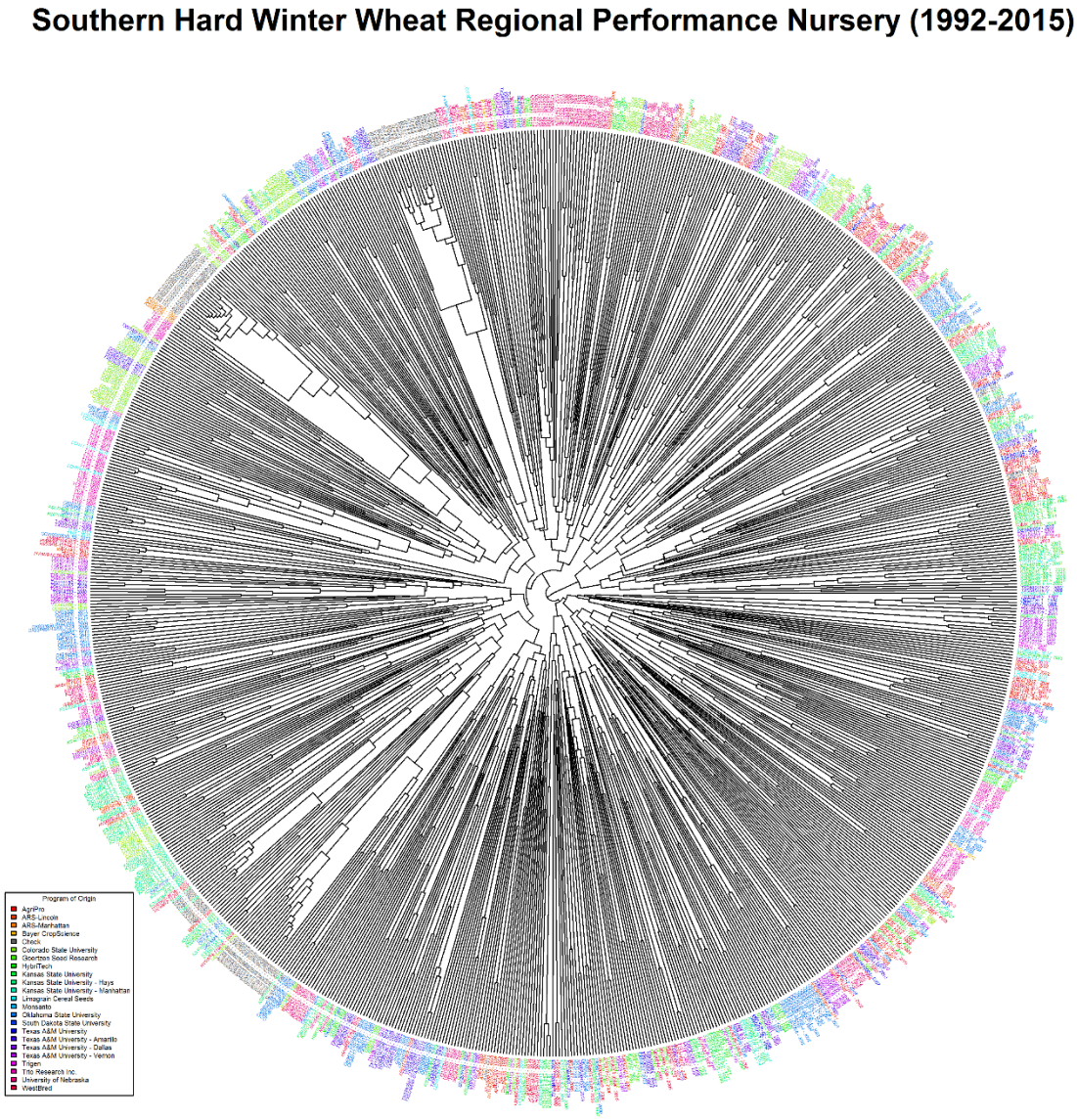

Figure 5-5. A dendrogram of the wheat lines used in this study created using the gbs.dendro function in the gbs-r package in R (unpublished). Color is used to group lines based on breeding program.

Appendix C - Supplementary Materials Chapter 4

Table C-1. Wheat varieties used in this analysis.

2180
Above
Akron
Alice
Anton
Arlin
Arrow
Avalanche
Baker's White
Bill Brown
Bison
Burchett
Caprock
Carson
Cheney
Clara CL (w)
CO04025
CO04393
CO04499
CO04W320
Comanche
Crest
Darrell
Doans
Dodge
Duke
Eagle
Expedition
Genou
Hail
Halt
Hatcher
HG-9
Hitch
Ike
JackPot
Jagalene
Jagger

Jerry
Jules
KARL_92
Kaw 61
Kiowa
Kirwin
Lakin
Lamar
Lancer
Larned
Lockett
Longhorn
Mace
McGill
Mit
MT0495
MT06103
MT9513
MT9904
MT9982
MTS0531
NE05496
Newton
NI08708
Norkan
OK06319
OK07209
OK07214
OK07S117
OK1068026
Parker
Parker 76
Prairie Red
Prowers
Ripper
Robidoux
Rosebud
Ruby Lee
Sage
Sandy
Scout 66

SD00111-9
SD01058
SD01237
SD05118
SD05210
SD05W018
Settler CL
Shawnee
Snowmass (w)
Stanton
Stout
Sturdy
Sturdy 2K
Sy Exp 1029
Sy Exp 38-45
Sy-Gold
Sy-Wolf
T-153
T-154
T-158
TAM 105
TAM 107
TAM 107-R7
TAM 109
TAM 110
TAM 111
TAM 112
TAM 113
TAM 200
TAM 202
TAM 203
TAM 302
TAM 303
TAM 304
TAM 401
TAM W-101
TAM400
Tascosa
Tiger
Trego
Trison

TX00V1131
TX01A5936
TX01V5134RC-3
TX02A0252
TX03A0148
TX04A001246
TX04M410164
TX04V075080
TX86A5606
TX86A6880
TX86A8072
TX99A0153-1
TX99U8618
Wendy
Wesley
Wichita
Windstar
Yellowstone
Yuma
Yumar
Aspen (W)
CO03064
Guymon
Judith
MT85200
NE05430
NE06545
Norris
NuSky
TX01M5009-28
TX03A0563
TX04M410211
TX96D1073

Table C-2. Loci, target alleles, and primer sequences used for sGBS.

locus_name	allele_a	allele_b	forward_primer	reverse_primer
BS00023148	CTCAAGGCTTTT	CTCAAGACTTTT	TGTAAAACGACGGCCAGTCCTCACTACAATGCAGCTCAAG	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATCTTTAGCCATCAAGATCCAGCACCAA
BS00067189	GCATGAATTAG	GCATGAATTAC	TGTAAAACGACGGCCAGTCTTATACAGGTAGACGCATGAATTA	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATCGCTTGCACAACTGCTTGTTCATGTA
BS00083385	GCGGTCTTCAGATGGT	GCGGTCTTCACATGGT	TGTAAAACGACGGCCAGTCAGCAGGTGGCGGTCTTCA	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATGGAGAAGTGCAGTGTCATCACCAT
BS00088726	ATACGAAGTATCATGGCGTATATGTAT	ATACGAAGTATCATGGCGTATATGTAC	TGTAAAACGACGGCCAGTATACGAAGTATCATGGCGTATATGTA	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATCGATGAATATTAGGTCTTACACATGTTCTT
BS00089969	TCTAGCTCCCTG	CTAGCCCCCTG	TGTAAAACGACGGCCAGTATAGCCGAAGCAGCTCTAGC	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATGTGCCGATAAGGAGAGCCCGTT
BS00150192	TAGATCAATTCATTCAG	TAGATCAACTCATTCAG	TGTAAAACGACGGCCAGTGAGAAGGGATGGAGATAGATCAA	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATCTCCCTCGGGTCTGGATTCTGAA
LR34exon11	TTCCATCATGATTATGTTAA	TTCCATCTTCATGATTATGTTAA	TGTAAAACGACGGCCAGTTTGCCATTATTGCACTCGTAAC	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATCCATGATGAATAGAAATAGTAGCTC
LR34exon11kasp	CTGGTATGCCATTTAACATAATCATGAA	CTGGTATGCCATTTAACATAATCATGAT	TGTAAAACGACGGCCAGTCTGGTATGCCATTTAACATAATCATGA	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATCGCATGACAATAAGTTTCACTCATGCAAA
Lr34exon12kasp	CGCAGTATCGA	CGCAGCATCGA	TGTAAAACGACGGCCAGTCATCATTCAGTCACCTCGCAG	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATGTGTTTGGAAGTATGAAGCAATAAATCGAT
Lr34exon22kasp	GAGATTTGCAGGAATG	GAGATTTGCATGAATG	TGTAAAACGACGGCCAGTTGTAATGTATCGTGAGAGATTTGCA	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATGATCATTATCTGACCTGTGCGAATGAATA
Lr34intron4kasp	TCCTCCGTCTTCTG	CCTCCGACTTCTG	TGTAAAACGACGGCCAGTACTCTTGCACAACCTCCTCCG	CCACTACGCCTCCGCTTTCCTCTCTCTATGGGCAGTCGGTGATTTGTGTCACCGGTGGCGCGTTT

Table C-3. Barcode sequences and forward oligo sequences.

set	well_A01	well_01A	well	barcode	oligo_sequence	adapter_name
spike_96A	A01	01A	1A	CGCGTGAACA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCGTGAACATGTAAAACGACGGCCAGT	ION_M13-384A_CGCGTGAACA_A01
spike_96A	B01	01B	1B	GCTTAGCGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTTAGCGGTTGTAAAACGACGGCCAGT	ION_M13-384A_GCTTAGCGGT_B01
spike_96A	C01	01C	1C	AGGATGCTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GGATGCTCTTGTAAAACGACGGCCAGT	ION_M13-384A_AGGATGCTCT_C01
spike_96A	D01	01D	1D	ATAACTGCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TAACTGCTTTGTAAAACGACGGCCAGT	ION_M13-384A_ATAACTGCTT_D01
spike_96A	E01	01E	1E	TTGGCTACGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGGCTACGTTGTAAAACGACGGCCAGT	ION_M13-384A_TTGGCTACGT_E01
spike_96A	F01	01F	1F	GTCAACTTAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCAACTTATTGTAAAACGACGGCCAGT	ION_M13-384A_GTCAACTTAT_F01
spike_96A	G01	01G	1G	GGCTCGAATG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GCTCGAATGTGTAAAACGACGGCCAGT	ION_M13-384A_GGCTCGAATG_G01
spike_96A	H01	01H	1H	TGCCTAATCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCCTAATCTTGTAAAACGACGGCCAGT	ION_M13-384A_TGCCTAATCT_H01
spike_96A	A02	02A	2A	GTTGCCTTCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TTGCCTTCATGTAAAACGACGGCCAGT	ION_M13-384A_GTTGCCTTCA_A02
spike_96A	B02	02B	2B	TGTTGCGTGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTTGCGTGCTGTAAAACGACGGCCAGT	ION_M13-384A_TGTTGCGTGC_B02
spike_96A	C02	02C	2C	TCGAGACCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CGAGACCTTTGTAAAACGACGGCCAGT	ION_M13-384A_TCGAGACCTT_C02
spike_96A	D02	02D	2D	ACAAGAATCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CAAGAATCGTGTAAAACGACGGCCAGT	ION_M13-384A_ACAAGAATCG_D02
spike_96A	E02	02E	2E	TGCACGGCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCACGGCATTGTAAAACGACGGCCAGT	ION_M13-384A_TGCACGGCAT_E02
spike_96A	F02	02F	2F	GGCGTCTCCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GCGTCTCCTTGTAAAACGACGGCCAGT	ION_M13-384A_GGCGTCTCCT_F02
spike_96A	G02	02G	2G	TGAGTTAGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GAGTTAGGCTGTAAAACGACGGCCAGT	ION_M13-384A_TGAGTTAGGC_G02

spike_96A	H02	02H	2H	CTCAGACAAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TCAGACAAGTGTAAAACGACGGCCAGT	ION_M13-384A_CTCAGACAAG_H02
spike_96A	A03	03A	3A	AGGTCAATTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GGTCAATTCTGTAAAACGACGGCCAGT	ION_M13-384A_AGGTCAATTC_A03
spike_96A	B03	03B	3B	AGCTTAGGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCTTAGGATTGTAAAACGACGGCCAGT	ION_M13-384A_AGCTTAGGAT_B03
spike_96A	C03	03C	3C	CGCGAGTGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCGAGTGCCTGTAAAACGACGGCCAGT	ION_M13-384A_CGCGAGTGCC_C03
spike_96A	D03	03D	3D	TTGTCGCATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGTCGCATTTGTAAAACGACGGCCAGT	ION_M13-384A_TTGTCGCATT_D03
spike_96A	E03	03E	3E	CAATGGTAAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AATGGTAACTGTAAAACGACGGCCAGT	ION_M13-384A_CAATGGTAAC_E03
spike_96A	F03	03F	3F	ATCACTCATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TCACTCATTTGTAAAACGACGGCCAGT	ION_M13-384A_ATCACTCATT_F03
spike_96A	G03	03G	3G	CGGCTAACTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GGCTAACTTTGTAAAACGACGGCCAGT	ION_M13-384A_CGGCTAACTT_G03
spike_96A	H03	03H	3H	CCAGTGGATC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CAGTGGATCTGTAAAACGACGGCCAGT	ION_M13-384A_CCAGTGGATC_H03
spike_96A	A04	04A	4A	TATTATCTAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ATTATCTAATGTAAAACGACGGCCAGT	ION_M13-384A_TATTATCTAA_A04
spike_96A	B04	04B	4B	GGCTAGGTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GCTAGGTGTTGTAAAACGACGGCCAGT	ION_M13-384A_GGCTAGGTGT_B04
spike_96A	C04	04C	4C	TGCTGCCACA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCTGCCACATGTAAAACGACGGCCAGT	ION_M13-384A_TGCTGCCACA_C04
spike_96A	D04	04D	4D	TTGCCGTCCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGCCGTCCTTGTAAAACGACGGCCAGT	ION_M13-384A_TTGCCGTCCT_D04
spike_96A	E04	04E	4E	AAGTACCTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA AGTACCTTATGTAAAACGACGGCCAGT	ION_M13-384A_AAGTACCTTA_E04
spike_96A	F04	04F	4F	TGGCCGCCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GGCCGCCTTTGTAAAACGACGGCCAGT	ION_M13-384A_TGGCCGCCTT_F04
spike_96A	G04	04G	4G	GCCGGAAGTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCGGAAGTATGTAAAACGACGGCCAGT	ION_M13-384A_GCCGGAAGTA_G04
spike_96A	H04	04H	4H	CCTTGACGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CTTGACGTTTGTAAAACGACGGCCAGT	ION_M13-384A_CCTTGACGTT_H04

spike_96A	A05	05A	5A	ACTCCTAGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTCCTAGATTGTAAAACGACGGCCAGT	ION_M13-384A_ACTCCTAGAT_A05
spike_96A	B05	05B	5B	CTTGACAGCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTGACAGCGTGTAAAACGACGGCCAGT	ION_M13-384A_CTTGACAGCG_B05
spike_96A	C05	05C	5C	CAGAGCTGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AGAGCTGCCTGTAAAACGACGGCCAGT	ION_M13-384A_CAGAGCTGCC_C05
spike_96A	D05	05D	5D	ATGCTTGAAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TGCTTGAATTGTAAAACGACGGCCAGT	ION_M13-384A_ATGCTTGAAT_D05
spike_96A	E05	05E	5E	CGCGCTAGAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCGCTAGAATGTAAAACGACGGCCAGT	ION_M13-384A_CGCGCTAGAA_E05
spike_96A	F05	05F	5F	CGCACGTCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCACGTCGTTGTAAAACGACGGCCAGT	ION_M13-384A_CGCACGTCGT_F05
spike_96A	G05	05G	5G	ATGCCACGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TGCCACGATTGTAAAACGACGGCCAGT	ION_M13-384A_ATGCCACGAT_G05
spike_96A	H05	05H	5H	GAATCCGAAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AATCCGAACTGTAAAACGACGGCCAGT	ION_M13-384A_GAATCCGAAC_H05
spike_96A	A06	06A	6A	AACGCGGAAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACGCGGAAGTGTAAAACGACGGCCAGT	ION_M13-384A_AACGCGGAAG_A06
spike_96A	B06	06B	6B	GTATCGAGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TATCGAGGCTGTAAAACGACGGCCAGT	ION_M13-384A_GTATCGAGGC_B06
spike_96A	C06	06C	6C	CTTACATAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTACATAGTTGTAAAACGACGGCCAGT	ION_M13-384A_CTTACATAGT_C06
spike_96A	D06	06D	6D	TGATGATCGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GATGATCGATGTAAAACGACGGCCAGT	ION_M13-384A_TGATGATCGA_D06
spike_96A	E06	06E	6E	ACACATCCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CACATCCGTTGTAAAACGACGGCCAGT	ION_M13-384A_ACACATCCGT_E06
spike_96A	F06	06F	6F	ACTTCATACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTTCATACCTGTAAAACGACGGCCAGT	ION_M13-384A_ACTTCATACC_F06
spike_96A	G06	06G	6G	CAATCTGACA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AATCTGACATGTAAAACGACGGCCAGT	ION_M13-384A_CAATCTGACA_G06
spike_96A	H06	06H	6H	GGATATAGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GATATAGGCTGTAAAACGACGGCCAGT	ION_M13-384A_GGATATAGGC_H06
spike_96A	A07	07A	7A	ACAATGCTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CAATGCTGATGTAAAACGACGGCCAGT	ION_M13-384A_ACAATGCTGA_A07

spike_96A	B07	07B	7B	GTCGGTAGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCGGTAGGTTGTAAAACGACGGCCAGT	ION_M13-384A_GTCGGTAGGT_B07
spike_96A	C07	07C	7C	TACGATTACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACGATTACTTGTAAAACGACGGCCAGT	ION_M13-384A_TACGATTACT_C07
spike_96A	D07	07D	7D	CGTCGATTGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTCGATTGCTGTAAAACGACGGCCAGT	ION_M13-384A_CGTCGATTGC_D07
spike_96A	E07	07E	7E	TAGCGCCAAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGCGCCAAGTGTAAAACGACGGCCAGT	ION_M13-384A_TAGCGCCAAG_E07
spike_96A	F07	07F	7F	TACGCATTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACGCATTGTTGTAAAACGACGGCCAGT	ION_M13-384A_TACGCATTGT_F07
spike_96A	G07	07G	7G	CAAGACATCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AAGACATCGTGTAAAACGACGGCCAGT	ION_M13-384A_CAAGACATCG_G07
spike_96A	H07	07H	7H	GAGTTAGAAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGTTAGAACTGTAAAACGACGGCCAGT	ION_M13-384A_GAGTTAGAAC_H07
spike_96A	A08	08A	8A	GCCTGCGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCTGCGATTTGTAAAACGACGGCCAGT	ION_M13-384A_GCCTGCGATT_A08
spike_96A	B08	08B	8B	TTGAGCTACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGAGCTACCTGTAAAACGACGGCCAGT	ION_M13-384A_TTGAGCTACC_B08
spike_96A	C08	08C	8C	TGCCTGCATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCCTGCATTTGTAAAACGACGGCCAGT	ION_M13-384A_TGCCTGCATT_C08
spike_96A	D08	08D	8D	CGCATAGTAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCATAGTAGTGTAAAACGACGGCCAGT	ION_M13-384A_CGCATAGTAG_D08
spike_96A	E08	08E	8E	CTTCTCACTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTCTCACTTTGTAAAACGACGGCCAGT	ION_M13-384A_CTTCTCACTT_E08
spike_96A	F08	08F	8F	GCTCCAGGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTCCAGGATTGTAAAACGACGGCCAGT	ION_M13-384A_GCTCCAGGAT_F08
spike_96A	G08	08G	8G	TACACGTGCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACACGTGCGTGTAAAACGACGGCCAGT	ION_M13-384A_TACACGTGCG_G08
spike_96A	H08	08H	8H	CAACGGCCAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AACGGCCACTGTAAAACGACGGCCAGT	ION_M13-384A_CAACGGCCAC_H08
spike_96A	A09	09A	9A	ACGTGTCCTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGTGTCCTGTGTAAAACGACGGCCAGT	ION_M13-384A_ACGTGTCCTG_A09
spike_96A	B09	09B	9B	TGGCGCACGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GGCGCACGTTGTAAAACGACGGCCAGT	ION_M13-384A_TGGCGCACGT_B09

spike_96A	C09	09C	9C	TTACTGCGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TACTGCGGCTGTAAAACGACGGCCAGT	ION_M13-384A_TTACTGCGGC_C09
spike_96A	D09	09D	9D	GTCCTCTCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCCTCTCGTTGTAAAACGACGGCCAGT	ION_M13-384A_GTCCTCTCGT_D09
spike_96A	E09	09E	9E	ATATGGCGTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TATGGCGTGTGTAAAACGACGGCCAGT	ION_M13-384A_ATATGGCGTG_E09
spike_96A	F09	09F	9F	AAGAATTAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA AGAATTAGTTGTAAAACGACGGCCAGT	ION_M13-384A_AAGAATTAGT_F09
spike_96A	G09	09G	9G	ACGCAGAAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGCAGAAGTTGTAAAACGACGGCCAGT	ION_M13-384A_ACGCAGAAGT_G09
spike_96A	H09	09H	9H	ACACGGCAGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CACGGCAGGTGTAAAACGACGGCCAGT	ION_M13-384A_ACACGGCAGG_H09
spike_96A	A10	10A	10A	GGACTATAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GACTATAGTTGTAAAACGACGGCCAGT	ION_M13-384A_GGACTATAGT_A10
spike_96A	B10	10B	10B	TCCTACGTAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CCTACGTACTGTAAAACGACGGCCAGT	ION_M13-384A_TCCTACGTAC_B10
spike_96A	C10	10C	10C	AGGAGGAGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GGAGGAGCTTGTAAAACGACGGCCAGT	ION_M13-384A_AGGAGGAGCT_C10
spike_96A	D10	10D	10D	TAGGAAGTAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGGAAGTAGTGTAAAACGACGGCCAGT	ION_M13-384A_TAGGAAGTAG_D10
spike_96A	E10	10E	10E	AACTGATTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACTGATTCCTGTAAAACGACGGCCAGT	ION_M13-384A_AACTGATTCC_E10
spike_96A	F10	10F	10F	GTAGGCTCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TAGGCTCTTTGTAAAACGACGGCCAGT	ION_M13-384A_GTAGGCTCTT_F10
spike_96A	G10	10G	10G	CTAGACCGTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TAGACCGTCTGTAAAACGACGGCCAGT	ION_M13-384A_CTAGACCGTC_G10
spike_96A	H10	10H	10H	CACGGCTTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACGGCTTCTTGTAAAACGACGGCCAGT	ION_M13-384A_CACGGCTTCT_H10
spike_96A	A11	11A	11A	TACACAAGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACACAAGCCTGTAAAACGACGGCCAGT	ION_M13-384A_TACACAAGCC_A11
spike_96A	B11	11B	11B	AAGTTCATAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA AGTTCATAATGTAAAACGACGGCCAGT	ION_M13-384A_AAGTTCATAA_B11
spike_96A	C11	11C	11C	TCTTACTCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTTACTCGCTGTAAAACGACGGCCAGT	ION_M13-384A_TCTTACTCGC_C11

spike_96A	D11	11D	11D	TCTACATCCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTACATCCGTGTAAAACGACGGCCAGT	ION_M13-384A_TCTACATCCG_D11
spike_96A	E11	11E	11E	GCCTCGTGGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCTCGTGGATGTAAAACGACGGCCAGT	ION_M13-384A_GCCTCGTGGA_E11
spike_96A	F11	11F	11F	CGTGTGCCGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTGTGCCGATGTAAAACGACGGCCAGT	ION_M13-384A_CGTGTGCCGA_F11
spike_96A	G11	11G	11G	TTGCATCGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGCATCGCCTGTAAAACGACGGCCAGT	ION_M13-384A_TTGCATCGCC_G11
spike_96A	H11	11H	11H	AACTACAACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACTACAACTTGTAAAACGACGGCCAGT	ION_M13-384A_AACTACAACT_H11
spike_96A	A12	12A	12A	TGCTACTTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCTACTTGATGTAAAACGACGGCCAGT	ION_M13-384A_TGCTACTTGA_A12
spike_96A	B12	12B	12B	CTCATTGACG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TCATTGACGTGTAAAACGACGGCCAGT	ION_M13-384A_CTCATTGACG_B12
spike_96A	C12	12C	12C	GGTGTACCGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTGTACCGATGTAAAACGACGGCCAGT	ION_M13-384A_GGTGTACCGA_C12
spike_96A	D12	12D	12D	CGTACTCGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTACTCGATTGTAAAACGACGGCCAGT	ION_M13-384A_CGTACTCGAT_D12
spike_96A	E12	12E	12E	GTGTACTAAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGTACTAATTGTAAAACGACGGCCAGT	ION_M13-384A_GTGTACTAAT_E12
spike_96A	F12	12F	12F	GGCTACACGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GCTACACGGTGTAAAACGACGGCCAGT	ION_M13-384A_GGCTACACGG_F12
spike_96A	G12	12G	12G	TGCTCAGTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCTCAGTTATGTAAAACGACGGCCAGT	ION_M13-384A_TGCTCAGTTA_G12
spike_96A	H12	12H	12H	ACATTCTAAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CATTCTAAGTGTAAAACGACGGCCAGT	ION_M13-384A_ACATTCTAAG_H12
spike_96B	A01	01A	1A	TCAGCGTCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CAGCGTCGTTGTAAAACGACGGCCAGT	ION_M13-384B_TCAGCGTCGT_A01
spike_96B	B01	01B	1B	CTTGGCGTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTGGCGTTATGTAAAACGACGGCCAGT	ION_M13-384B_CTTGGCGTTA_B01
spike_96B	C01	01C	1C	AGACCATTAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GACCATTAGTGTAAAACGACGGCCAGT	ION_M13-384B_AGACCATTAG_C01
spike_96B	D01	01D	1D	ACAGTAATCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CAGTAATCGTGTAAAACGACGGCCAGT	ION_M13-384B_ACAGTAATCG_D01

spike_96B	E01	01E	1E	ACTCAATTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTCAATTGATGTAAAACGACGGCCAGT	ION_M13-384B_ACTCAATTGA_E01
spike_96B	F01	01F	1F	AGCCACAGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCCACAGCTTGTAAAACGACGGCCAGT	ION_M13-384B_AGCCACAGCT_F01
spike_96B	G01	01G	1G	GCATTAGCAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CATTAGCACTGTAAAACGACGGCCAGT	ION_M13-384B_GCATTAGCAC_G01
spike_96B	H01	01H	1H	AGGTGGTTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GGTGGTTGATGTAAAACGACGGCCAGT	ION_M13-384B_AGGTGGTTGA_H01
spike_96B	A02	02A	2A	AATCGTATCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ATCGTATCTTGTAAAACGACGGCCAGT	ION_M13-384B_AATCGTATCT_A02
spike_96B	B02	02B	2B	GTTCCACTGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TTCCACTGGTGTAAAACGACGGCCAGT	ION_M13-384B_GTTCCACTGG_B02
spike_96B	C02	02C	2C	CGCCAGAGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCCAGAGTTTGTAAAACGACGGCCAGT	ION_M13-384B_CGCCAGAGTT_C02
spike_96B	D02	02D	2D	CTTGTGGTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTGTGGTCTTGTAAAACGACGGCCAGT	ION_M13-384B_CTTGTGGTCT_D02
spike_96B	E02	02E	2E	GTCCGTCTGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCCGTCTGCTGTAAAACGACGGCCAGT	ION_M13-384B_GTCCGTCTGC_E02
spike_96B	F02	02F	2F	GTATTATAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TATTATAGTTGTAAAACGACGGCCAGT	ION_M13-384B_GTATTATAGT_F02
spike_96B	G02	02G	2G	TCCTTATGAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CCTTATGAATGTAAAACGACGGCCAGT	ION_M13-384B_TCCTTATGAA_G02
spike_96B	H02	02H	2H	AGTAACGCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTAACGCATTGTAAAACGACGGCCAGT	ION_M13-384B_AGTAACGCAT_H02
spike_96B	A03	03A	3A	CACTCGAGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACTCGAGGTTGTAAAACGACGGCCAGT	ION_M13-384B_CACTCGAGGT_A03
spike_96B	B03	03B	3B	CCTAGAGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CTAGAGATTTGTAAAACGACGGCCAGT	ION_M13-384B_CCTAGAGATT_B03
spike_96B	C03	03C	3C	GCGCTGCTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CGCTGCTGATGTAAAACGACGGCCAGT	ION_M13-384B_GCGCTGCTGA_C03
spike_96B	D03	03D	3D	TTCTATTCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TCTATTCGCTGTAAAACGACGGCCAGT	ION_M13-384B_TTCTATTCGC_D03
spike_96B	E03	03E	3E	AGCACAGCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCACAGCGCTGTAAAACGACGGCCAGT	ION_M13-384B_AGCACAGCGC_E03

spike_96B	F03	03F	3F	TTAGTTCATA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TAGTTCATATGTAAAACGACGGCCAGT	ION_M13-384B_TTAGTTCATA_F03
spike_96B	G03	03G	3G	TCCACCGCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CCACCGCTCTGTAAAACGACGGCCAGT	ION_M13-384B_TCCACCGCTC_G03
spike_96B	H03	03H	3H	CCATATGCGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CATATGCGGTGTAAAACGACGGCCAGT	ION_M13-384B_CCATATGCGG_H03
spike_96B	A04	04A	4A	GACTAAGACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACTAAGACTTGTAAAACGACGGCCAGT	ION_M13-384B_GACTAAGACT_A04
spike_96B	B04	04B	4B	CTCGTTATGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TCGTTATGCTGTAAAACGACGGCCAGT	ION_M13-384B_CTCGTTATGC_B04
spike_96B	C04	04C	4C	CTTCTATAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTCTATAGTTGTAAAACGACGGCCAGT	ION_M13-384B_CTTCTATAGT_C04
spike_96B	D04	04D	4D	CGTGGTCAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTGGTCAGTTGTAAAACGACGGCCAGT	ION_M13-384B_CGTGGTCAGT_D04
spike_96B	E04	04E	4E	TAGGTGAATG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGGTGAATGTGTAAAACGACGGCCAGT	ION_M13-384B_TAGGTGAATG_E04
spike_96B	F04	04F	4F	AGTATAAGTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTATAAGTCTGTAAAACGACGGCCAGT	ION_M13-384B_AGTATAAGTC_F04
spike_96B	G04	04G	4G	GCCACGCTAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCACGCTAATGTAAAACGACGGCCAGT	ION_M13-384B_GCCACGCTAA_G04
spike_96B	H04	04H	4H	TCCTCCAGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CCTCCAGGTTGTAAAACGACGGCCAGT	ION_M13-384B_TCCTCCAGGT_H04
spike_96B	A05	05A	5A	TGATTCATCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GATTCATCCTGTAAAACGACGGCCAGT	ION_M13-384B_TGATTCATCC_A05
spike_96B	B05	05B	5B	GACGAGACGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACGAGACGATGTAAAACGACGGCCAGT	ION_M13-384B_GACGAGACGA_B05
spike_96B	C05	05C	5C	CACTACTTAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACTACTTAATGTAAAACGACGGCCAGT	ION_M13-384B_CACTACTTAA_C05
spike_96B	D05	05D	5D	AGAGTGTAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GAGTGTAGTTGTAAAACGACGGCCAGT	ION_M13-384B_AGAGTGTAGT_D05
spike_96B	E05	05E	5E	CTGCGGAGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TGCGGAGGTTGTAAAACGACGGCCAGT	ION_M13-384B_CTGCGGAGGT_E05
spike_96B	F05	05F	5F	GGTCCTCAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTCCTCAGTTGTAAAACGACGGCCAGT	ION_M13-384B_GGTCCTCAGT_F05

spike_96B	G05	05G	5G	GGTGTCAGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTGTCAGTTTGTAAAACGACGGCCAGT	ION_M13-384B_GGTGTCAGTT_G05
spike_96B	H05	05H	5H	GTTCGATCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TTCGATCATTGTAAAACGACGGCCAGT	ION_M13-384B_GTTCGATCAT_H05
spike_96B	A06	06A	6A	TTCAACGCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TCAACGCTTTGTAAAACGACGGCCAGT	ION_M13-384B_TTCAACGCTT_A06
spike_96B	B06	06B	6B	GATGGTAGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATGGTAGTTTGTAAAACGACGGCCAGT	ION_M13-384B_GATGGTAGTT_B06
spike_96B	C06	06C	6C	TACCGAACGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACCGAACGTTGTAAAACGACGGCCAGT	ION_M13-384B_TACCGAACGT_C06
spike_96B	D06	06D	6D	AGGCGACCAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GGCGACCACTGTAAAACGACGGCCAGT	ION_M13-384B_AGGCGACCAC_D06
spike_96B	E06	06E	6E	TCGCACTTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CGCACTTGTTGTAAAACGACGGCCAGT	ION_M13-384B_TCGCACTTGT_E06
spike_96B	F06	06F	6F	ATCATACCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TCATACCTCTGTAAAACGACGGCCAGT	ION_M13-384B_ATCATACCTC_F06
spike_96B	G06	06G	6G	CAACTAACAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AACTAACATTGTAAAACGACGGCCAGT	ION_M13-384B_CAACTAACAT_G06
spike_96B	H06	06H	6H	GACCAGCCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACCAGCCATTGTAAAACGACGGCCAGT	ION_M13-384B_GACCAGCCAT_H06
spike_96B	A07	07A	7A	GCATTGTGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CATTGTGTTTGTAAAACGACGGCCAGT	ION_M13-384B_GCATTGTGTT_A07
spike_96B	B07	07B	7B	GCGTGCACTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CGTGCACTGTGTAAAACGACGGCCAGT	ION_M13-384B_GCGTGCACTG_B07
spike_96B	C07	07C	7C	TGATCCTACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GATCCTACCTGTAAAACGACGGCCAGT	ION_M13-384B_TGATCCTACC_C07
spike_96B	D07	07D	7D	ACTTAACAAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTTAACAATTGTAAAACGACGGCCAGT	ION_M13-384B_ACTTAACAAT_D07
spike_96B	E07	07E	7E	TGTGAGCTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTGAGCTCCTGTAAAACGACGGCCAGT	ION_M13-384B_TGTGAGCTCC_E07
spike_96B	F07	07F	7F	AACAGCGAAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACAGCGAAGTGTAAAACGACGGCCAGT	ION_M13-384B_AACAGCGAAG_F07
spike_96B	G07	07G	7G	GTTATCCGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TTATCCGCTTGTAAAACGACGGCCAGT	ION_M13-384B_GTTATCCGCT_G07

spike_96B	H07	07H	7H	CGATCATGAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GATCATGAATGTAAAACGACGGCCAGT	ION_M13-384B_CGATCATGAA_H07
spike_96B	A08	08A	8A	CGCAGGCTAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCAGGCTAATGTAAAACGACGGCCAGT	ION_M13-384B_CGCAGGCTAA_A08
spike_96B	B08	08B	8B	CATCAGAGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ATCAGAGCTTGTAAAACGACGGCCAGT	ION_M13-384B_CATCAGAGCT_B08
spike_96B	C08	08C	8C	GAGTGATGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGTGATGGCTGTAAAACGACGGCCAGT	ION_M13-384B_GAGTGATGGC_C08
spike_96B	D08	08D	8D	CGAGTTGCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GAGTTGCGCTGTAAAACGACGGCCAGT	ION_M13-384B_CGAGTTGCGC_D08
spike_96B	E08	08E	8E	GGTAGCTACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTAGCTACCTGTAAAACGACGGCCAGT	ION_M13-384B_GGTAGCTACC_E08
spike_96B	F08	08F	8F	GTTGGAGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TTGGAGATTTGTAAAACGACGGCCAGT	ION_M13-384B_GTTGGAGATT_F08
spike_96B	G08	08G	8G	AGTGGAGGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTGGAGGTTTGTAAAACGACGGCCAGT	ION_M13-384B_AGTGGAGGTT_G08
spike_96B	H08	08H	8H	GTGGTGGTAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGGTGGTATTGTAAAACGACGGCCAGT	ION_M13-384B_GTGGTGGTAT_H08
spike_96B	A09	09A	9A	GTGATAGCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGATAGCGTTGTAAAACGACGGCCAGT	ION_M13-384B_GTGATAGCGT_A09
spike_96B	B09	09B	9B	GTCTCTACGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCTCTACGTTGTAAAACGACGGCCAGT	ION_M13-384B_GTCTCTACGT_B09
spike_96B	C09	09C	9C	AGCCTTGGTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCCTTGGTATGTAAAACGACGGCCAGT	ION_M13-384B_AGCCTTGGTA_C09
spike_96B	D09	09D	9D	CGACCGTCGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GACCGTCGGTGTAAAACGACGGCCAGT	ION_M13-384B_CGACCGTCGG_D09
spike_96B	E09	09E	9E	GGCTGTGTAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GCTGTGTAGTGTAAAACGACGGCCAGT	ION_M13-384B_GGCTGTGTAG_E09
spike_96B	F09	09F	9F	AGGAACTCCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GGAACTCCATGTAAAACGACGGCCAGT	ION_M13-384B_AGGAACTCCA_F09
spike_96B	G09	09G	9G	CCGTCGTCTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGTCGTCTGTGTAAAACGACGGCCAGT	ION_M13-384B_CCGTCGTCTG_G09
spike_96B	H09	09H	9H	AATCCACGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ATCCACGCCTGTAAAACGACGGCCAGT	ION_M13-384B_AATCCACGCC_H09

spike_96B	A10	10A	10A	ATTCGTTCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TTCGTTCTTTGTAAAACGACGGCCAGT	ION_M13-384B_ATTCGTTCTT_A10
spike_96B	B10	10B	10B	GTAGGACAGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TAGGACAGATGTAAAACGACGGCCAGT	ION_M13-384B_GTAGGACAGA_B10
spike_96B	C10	10C	10C	TGCTCGCTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCTCGCTCTTGTAAAACGACGGCCAGT	ION_M13-384B_TGCTCGCTCT_C10
spike_96B	D10	10D	10D	CCGGAAGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGGAAGATTTGTAAAACGACGGCCAGT	ION_M13-384B_CCGGAAGATT_D10
spike_96B	E10	10E	10E	TTCGAGGATC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TCGAGGATCTGTAAAACGACGGCCAGT	ION_M13-384B_TTCGAGGATC_E10
spike_96B	F10	10F	10F	GACACGGTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACACGGTTATGTAAAACGACGGCCAGT	ION_M13-384B_GACACGGTTA_F10
spike_96B	G10	10G	10G	ATATAGAACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TATAGAACCTGTAAAACGACGGCCAGT	ION_M13-384B_ATATAGAACC_G10
spike_96B	H10	10H	10H	AGCTAGTGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCTAGTGCATGTAAAACGACGGCCAGT	ION_M13-384B_AGCTAGTGCA_H10
spike_96B	A11	11A	11A	GTGGCGCTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGGCGCTGTTGTAAAACGACGGCCAGT	ION_M13-384B_GTGGCGCTGT_A11
spike_96B	B11	11B	11B	TGTACCTGAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTACCTGAGTGTAAAACGACGGCCAGT	ION_M13-384B_TGTACCTGAG_B11
spike_96B	C11	11C	11C	GTCGTCGTCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCGTCGTCATGTAAAACGACGGCCAGT	ION_M13-384B_GTCGTCGTCA_C11
spike_96B	D11	11D	11D	ACGAAGCTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGAAGCTTATGTAAAACGACGGCCAGT	ION_M13-384B_ACGAAGCTTA_D11
spike_96B	E11	11E	11E	CCTCAAGAAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CTCAAGAACTGTAAAACGACGGCCAGT	ION_M13-384B_CCTCAAGAAC_E11
spike_96B	F11	11F	11F	TGTGACTTAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTGACTTAGTGTAAAACGACGGCCAGT	ION_M13-384B_TGTGACTTAG_F11
spike_96B	G11	11G	11G	GATTCAATAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATTCAATAGTGTAAAACGACGGCCAGT	ION_M13-384B_GATTCAATAG_G11
spike_96B	H11	11H	11H	GTGGACGATA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGGACGATATGTAAAACGACGGCCAGT	ION_M13-384B_GTGGACGATA_H11
spike_96B	A12	12A	12A	ACGTGAAGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGTGAAGGCTGTAAAACGACGGCCAGT	ION_M13-384B_ACGTGAAGGC_A12

spike_96B	B12	12B	12B	CTAGCGCTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TAGCGCTCTTGTAAAACGACGGCCAGT	ION_M13-384B_CTAGCGCTCT_B12
spike_96B	C12	12C	12C	CCGCGATGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGCGATGTTTGTAAAACGACGGCCAGT	ION_M13-384B_CCGCGATGTT_C12
spike_96B	D12	12D	12D	CACTATGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACTATGATTTGTAAAACGACGGCCAGT	ION_M13-384B_CACTATGATT_D12
spike_96B	E12	12E	12E	ATAGGCGAGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TAGGCGAGGTGTAAAACGACGGCCAGT	ION_M13-384B_ATAGGCGAGG_E12
spike_96B	F12	12F	12F	ATAATAGTAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TAATAGTATTGTAAAACGACGGCCAGT	ION_M13-384B_ATAATAGTAT_F12
spike_96B	G12	12G	12G	TGGTAAGCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GGTAAGCGCTGTAAAACGACGGCCAGT	ION_M13-384B_TGGTAAGCGC_G12
spike_96B	H12	12H	12H	AGAGCAGGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GAGCAGGCTTGTAAAACGACGGCCAGT	ION_M13-384B_AGAGCAGGCT_H12
spike_96C	A01	01A	1A	CCAACTTAGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CAACTTAGATGTAAAACGACGGCCAGT	ION_M13-384C_CCAACTTAGA_A01
spike_96C	B01	01B	1B	TCGAATCCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CGAATCCTCTGTAAAACGACGGCCAGT	ION_M13-384C_TCGAATCCTC_B01
spike_96C	C01	01C	1C	TAATAGTGAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AATAGTGACTGTAAAACGACGGCCAGT	ION_M13-384C_TAATAGTGAC_C01
spike_96C	D01	01D	1D	CAAGCTCGTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AAGCTCGTCTGTAAAACGACGGCCAGT	ION_M13-384C_CAAGCTCGTC_D01
spike_96C	E01	01E	1E	CTGGCTGTCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TGGCTGTCGTGTAAAACGACGGCCAGT	ION_M13-384C_CTGGCTGTCG_E01
spike_96C	F01	01F	1F	GCCGCTCGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCGCTCGGTTGTAAAACGACGGCCAGT	ION_M13-384C_GCCGCTCGGT_F01
spike_96C	G01	01G	1G	CACGTGCACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACGTGCACTTGTAAAACGACGGCCAGT	ION_M13-384C_CACGTGCACT_G01
spike_96C	H01	01H	1H	GAGATGCAAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGATGCAATTGTAAAACGACGGCCAGT	ION_M13-384C_GAGATGCAAT_H01
spike_96C	A02	02A	2A	CGGACGAGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GGACGAGCATGTAAAACGACGGCCAGT	ION_M13-384C_CGGACGAGCA_A02
spike_96C	B02	02B	2B	CTGAGATGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TGAGATGATTGTAAAACGACGGCCAGT	ION_M13-384C_CTGAGATGAT_B02

spike_96C	C02	02C	2C	ACAACCGCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CAACCGCGTTGTAAAACGACGGCCAGT	ION_M13-384C_ACAACCGCGT_C02
spike_96C	D02	02D	2D	CGGCTCTCGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GGCTCTCGGTGTAAAACGACGGCCAGT	ION_M13-384C_CGGCTCTCGG_D02
spike_96C	E02	02E	2E	GTCAGAGTAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCAGAGTACTGTAAAACGACGGCCAGT	ION_M13-384C_GTCAGAGTAC_E02
spike_96C	F02	02F	2F	GGAGTCGATA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GAGTCGATATGTAAAACGACGGCCAGT	ION_M13-384C_GGAGTCGATA_F02
spike_96C	G02	02G	2G	GGAGGTGTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GAGGTGTTATGTAAAACGACGGCCAGT	ION_M13-384C_GGAGGTGTTA_G02
spike_96C	H02	02H	2H	TAGCATTGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGCATTGCTTGTAAAACGACGGCCAGT	ION_M13-384C_TAGCATTGCT_H02
spike_96C	A03	03A	3A	TCGAAGGATC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CGAAGGATCTGTAAAACGACGGCCAGT	ION_M13-384C_TCGAAGGATC_A03
spike_96C	B03	03B	3B	GAACGTAGGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AACGTAGGATGTAAAACGACGGCCAGT	ION_M13-384C_GAACGTAGGA_B03
spike_96C	C03	03C	3C	CTGGATAAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TGGATAAGTTGTAAAACGACGGCCAGT	ION_M13-384C_CTGGATAAGT_C03
spike_96C	D03	03D	3D	TATACACCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ATACACCATTGTAAAACGACGGCCAGT	ION_M13-384C_TATACACCAT_D03
spike_96C	E03	03E	3E	ATAAGTTCTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TAAGTTCTGTGTAAAACGACGGCCAGT	ION_M13-384C_ATAAGTTCTG_E03
spike_96C	F03	03F	3F	CGTGGCTTCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTGGCTTCGTGTAAAACGACGGCCAGT	ION_M13-384C_CGTGGCTTCG_F03
spike_96C	G03	03G	3G	CATCGGTGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ATCGGTGATTGTAAAACGACGGCCAGT	ION_M13-384C_CATCGGTGAT_G03
spike_96C	H03	03H	3H	GCTTGATCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTTGATCATTGTAAAACGACGGCCAGT	ION_M13-384C_GCTTGATCAT_H03
spike_96C	A04	04A	4A	TGACGAACTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GACGAACTATGTAAAACGACGGCCAGT	ION_M13-384C_TGACGAACTA_A04
spike_96C	B04	04B	4B	GCTGGCGGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTGGCGGTTTGTAAAACGACGGCCAGT	ION_M13-384C_GCTGGCGGTT_B04
spike_96C	C04	04C	4C	GTGATTAGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGATTAGATTGTAAAACGACGGCCAGT	ION_M13-384C_GTGATTAGAT_C04

spike_96C	D04	04D	4D	TTAACAGCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TAACAGCGTTGTAAAACGACGGCCAGT	ION_M13-384C_TTAACAGCGT_D04
spike_96C	E04	04E	4E	GAGAGTACGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGAGTACGGTGTAAAACGACGGCCAGT	ION_M13-384C_GAGAGTACGG_E04
spike_96C	F04	04F	4F	TTACTAGCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TACTAGCTTTGTAAAACGACGGCCAGT	ION_M13-384C_TTACTAGCTT_F04
spike_96C	G04	04G	4G	GCACGTTGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CACGTTGATTGTAAAACGACGGCCAGT	ION_M13-384C_GCACGTTGAT_G04
spike_96C	H04	04H	4H	AGCCTACCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCCTACCTCTGTAAAACGACGGCCAGT	ION_M13-384C_AGCCTACCTC_H04
spike_96C	A05	05A	5A	ATGAGAATCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TGAGAATCATGTAAAACGACGGCCAGT	ION_M13-384C_ATGAGAATCA_A05
spike_96C	B05	05B	5B	AGAGAGCCAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GAGAGCCAATGTAAAACGACGGCCAGT	ION_M13-384C_AGAGAGCCAA_B05
spike_96C	C05	05C	5C	AATATATGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ATATATGCATGTAAAACGACGGCCAGT	ION_M13-384C_AATATATGCA_C05
spike_96C	D05	05D	5D	CCTTCCAGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CTTCCAGGCTGTAAAACGACGGCCAGT	ION_M13-384C_CCTTCCAGGC_D05
spike_96C	E05	05E	5E	CAAGGAGCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AAGGAGCGCTGTAAAACGACGGCCAGT	ION_M13-384C_CAAGGAGCGC_E05
spike_96C	F05	05F	5F	TCCATGCCAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CCATGCCAGTGTAAAACGACGGCCAGT	ION_M13-384C_TCCATGCCAG_F05
spike_96C	G05	05G	5G	AGTCATCCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTCATCCGTTGTAAAACGACGGCCAGT	ION_M13-384C_AGTCATCCGT_G05
spike_96C	H05	05H	5H	TCAGGTCTGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CAGGTCTGCTGTAAAACGACGGCCAGT	ION_M13-384C_TCAGGTCTGC_H05
spike_96C	A06	06A	6A	AGTACGCTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTACGCTGTTGTAAAACGACGGCCAGT	ION_M13-384C_AGTACGCTGT_A06
spike_96C	B06	06B	6B	GATGATTCCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATGATTCCTTGTAAAACGACGGCCAGT	ION_M13-384C_GATGATTCCT_B06
spike_96C	C06	06C	6C	GGCGATGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GCGATGATTTGTAAAACGACGGCCAGT	ION_M13-384C_GGCGATGATT_C06
spike_96C	D06	06D	6D	AGTCGCTGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTCGCTGCATGTAAAACGACGGCCAGT	ION_M13-384C_AGTCGCTGCA_D06

spike_96C	E06	06E	6E	TGTGCCGCCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTGCCGCCTTGTAAAACGACGGCCAGT	ION_M13-384C_TGTGCCGCCT_E06
spike_96C	F06	06F	6F	TCTTGCAGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTTGCAGCCTGTAAAACGACGGCCAGT	ION_M13-384C_TCTTGCAGCC_F06
spike_96C	G06	06G	6G	ACGATAGATA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGATAGATATGTAAAACGACGGCCAGT	ION_M13-384C_ACGATAGATA_G06
spike_96C	H06	06H	6H	GGTTGACGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTTGACGATTGTAAAACGACGGCCAGT	ION_M13-384C_GGTTGACGAT_H06
spike_96C	A07	07A	7A	CCGTACGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGTACGATTTGTAAAACGACGGCCAGT	ION_M13-384C_CCGTACGATT_A07
spike_96C	B07	07B	7B	GTGGTCAAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGGTCAAGTTGTAAAACGACGGCCAGT	ION_M13-384C_GTGGTCAAGT_B07
spike_96C	C07	07C	7C	TCGCAAGTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CGCAAGTTATGTAAAACGACGGCCAGT	ION_M13-384C_TCGCAAGTTA_C07
spike_96C	D07	07D	7D	CAGCGTCCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AGCGTCCGTTGTAAAACGACGGCCAGT	ION_M13-384C_CAGCGTCCGT_D07
spike_96C	E07	07E	7E	TATCCGTAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ATCCGTAGTTGTAAAACGACGGCCAGT	ION_M13-384C_TATCCGTAGT_E07
spike_96C	F07	07F	7F	CAACCAGAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AACCAGAGTTGTAAAACGACGGCCAGT	ION_M13-384C_CAACCAGAGT_F07
spike_96C	G07	07G	7G	CAAGAATCAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AAGAATCACTGTAAAACGACGGCCAGT	ION_M13-384C_CAAGAATCAC_G07
spike_96C	H07	07H	7H	CGAGCCGAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GAGCCGAGTTGTAAAACGACGGCCAGT	ION_M13-384C_CGAGCCGAGT_H07
spike_96C	A08	08A	8A	AACCTAAGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACCTAAGCTTGTAAAACGACGGCCAGT	ION_M13-384C_AACCTAAGCT_A08
spike_96C	B08	08B	8B	AATGGCCATC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ATGGCCATCTGTAAAACGACGGCCAGT	ION_M13-384C_AATGGCCATC_B08
spike_96C	C08	08C	8C	TACATCACGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACATCACGGTGTAAAACGACGGCCAGT	ION_M13-384C_TACATCACGG_C08
spike_96C	D08	08D	8D	AACACACCAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACACACCAGTGTAAAACGACGGCCAGT	ION_M13-384C_AACACACCAG_D08
spike_96C	E08	08E	8E	GACTGCTTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACTGCTTGTTGTAAAACGACGGCCAGT	ION_M13-384C_GACTGCTTGT_E08

spike_96C	F08	08F	8F	GGATACGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GATACGATTTGTAAAACGACGGCCAGT	ION_M13-384C_GGATACGATT_F08
spike_96C	G08	08G	8G	CATCGAAGTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ATCGAAGTATGTAAAACGACGGCCAGT	ION_M13-384C_CATCGAAGTA_G08
spike_96C	H08	08H	8H	CGTCGTAATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTCGTAATTTGTAAAACGACGGCCAGT	ION_M13-384C_CGTCGTAATT_H08
spike_96C	A09	09A	9A	GGTGATCGAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTGATCGAGTGTAAAACGACGGCCAGT	ION_M13-384C_GGTGATCGAG_A09
spike_96C	B09	09B	9B	CTCAACAGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TCAACAGCCTGTAAAACGACGGCCAGT	ION_M13-384C_CTCAACAGCC_B09
spike_96C	C09	09C	9C	GATTCTGCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATTCTGCTTTGTAAAACGACGGCCAGT	ION_M13-384C_GATTCTGCTT_C09
spike_96C	D09	09D	9D	CACCGCGACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACCGCGACCTGTAAAACGACGGCCAGT	ION_M13-384C_CACCGCGACC_D09
spike_96C	E09	09E	9E	CACCTTCAGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACCTTCAGCTGTAAAACGACGGCCAGT	ION_M13-384C_CACCTTCAGC_E09
spike_96C	F09	09F	9F	GCAGCACGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CAGCACGCTTGTAAAACGACGGCCAGT	ION_M13-384C_GCAGCACGCT_F09
spike_96C	G09	09G	9G	TATCGATGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ATCGATGGTTGTAAAACGACGGCCAGT	ION_M13-384C_TATCGATGGT_G09
spike_96C	H09	09H	9H	GAGAATCATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGAATCATTTGTAAAACGACGGCCAGT	ION_M13-384C_GAGAATCATT_H09
spike_96C	A10	10A	10A	AACCTCCGAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACCTCCGAGTGTAAAACGACGGCCAGT	ION_M13-384C_AACCTCCGAG_A10
spike_96C	B10	10B	10B	TAACGGAGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AACGGAGCTTGTAAAACGACGGCCAGT	ION_M13-384C_TAACGGAGCT_B10
spike_96C	C10	10C	10C	CATTGTTCTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ATTGTTCTATGTAAAACGACGGCCAGT	ION_M13-384C_CATTGTTCTA_C10
spike_96C	D10	10D	10D	GCAAGCCGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CAAGCCGTTTGTAAAACGACGGCCAGT	ION_M13-384C_GCAAGCCGTT_D10
spike_96C	E10	10E	10E	CTCTATCGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TCTATCGATTGTAAAACGACGGCCAGT	ION_M13-384C_CTCTATCGAT_E10
spike_96C	F10	10F	10F	GCAACTATCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CAACTATCATGTAAAACGACGGCCAGT	ION_M13-384C_GCAACTATCA_F10

spike_96C	G10	10G	10G	CGTGCTTGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GTGCTTGATTGTAAAACGACGGCCAGT	ION_M13-384C_CGTGCTTGAT_G10
spike_96C	H10	10H	10H	GAAGCGAACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AAGCGAACTTGTAAAACGACGGCCAGT	ION_M13-384C_GAAGCGAACT_H10
spike_96C	A11	11A	11A	GTATGTATAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TATGTATAATGTAAAACGACGGCCAGT	ION_M13-384C_GTATGTATAA_A11
spike_96C	B11	11B	11B	GTCTCAGCTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCTCAGCTATGTAAAACGACGGCCAGT	ION_M13-384C_GTCTCAGCTA_B11
spike_96C	C11	11C	11C	GAGTAGCGTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGTAGCGTCTGTAAAACGACGGCCAGT	ION_M13-384C_GAGTAGCGTC_C11
spike_96C	D11	11D	11D	CACAAGCTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACAAGCTCTTGTAAAACGACGGCCAGT	ION_M13-384C_CACAAGCTCT_D11
spike_96C	E11	11E	11E	CTGTTAGGAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TGTTAGGACTGTAAAACGACGGCCAGT	ION_M13-384C_CTGTTAGGAC_E11
spike_96C	F11	11F	11F	TGCAGATGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCAGATGTTTGTAAAACGACGGCCAGT	ION_M13-384C_TGCAGATGTT_F11
spike_96C	G11	11G	11G	CACGAAGATA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACGAAGATATGTAAAACGACGGCCAGT	ION_M13-384C_CACGAAGATA_G11
spike_96C	H11	11H	11H	CCTATTGAGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CTATTGAGCTGTAAAACGACGGCCAGT	ION_M13-384C_CCTATTGAGC_H11
spike_96C	A12	12A	12A	ACCATTCTGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CCATTCTGCTGTAAAACGACGGCCAGT	ION_M13-384C_ACCATTCTGC_A12
spike_96C	B12	12B	12B	GAAGACTGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AAGACTGCCTGTAAAACGACGGCCAGT	ION_M13-384C_GAAGACTGCC_B12
spike_96C	C12	12C	12C	TCCGGCGCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CCGGCGCATTGTAAAACGACGGCCAGT	ION_M13-384C_TCCGGCGCAT_C12
spike_96C	D12	12D	12D	TTCTGGACAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TCTGGACAGTGTAAAACGACGGCCAGT	ION_M13-384C_TTCTGGACAG_D12
spike_96C	E12	12E	12E	GCGGTTCGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CGGTTCGATTGTAAAACGACGGCCAGT	ION_M13-384C_GCGGTTCGAT_E12
spike_96C	F12	12F	12F	GTAGTCCGGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TAGTCCGGTTGTAAAACGACGGCCAGT	ION_M13-384C_GTAGTCCGGT_F12
spike_96C	G12	12G	12G	GCCTCACGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCTCACGCCTGTAAAACGACGGCCAGT	ION_M13-384C_GCCTCACGCC_G12

spike_96C	H12	12H	12H	GTCATCATGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TCATCATGCTGTAAAACGACGGCCAGT	ION_M13-384C_GTCATCATGC_H12
spike_96D	A01	01A	1A	AATCTAGGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ATCTAGGTTTGTAAAACGACGGCCAGT	ION_M13-384D_AATCTAGGTT_A01
spike_96D	B01	01B	1B	TGTTGTCGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTTGTCGATTGTAAAACGACGGCCAGT	ION_M13-384D_TGTTGTCGAT_B01
spike_96D	C01	01C	1C	GTAGTGTTCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TAGTGTTCATGTAAAACGACGGCCAGT	ION_M13-384D_GTAGTGTTCA_C01
spike_96D	D01	01D	1D	ACTCCGTCCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTCCGTCCTTGTAAAACGACGGCCAGT	ION_M13-384D_ACTCCGTCCT_D01
spike_96D	E01	01E	1E	CGCGTATACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GCGTATACTTGTAAAACGACGGCCAGT	ION_M13-384D_CGCGTATACT_E01
spike_96D	F01	01F	1F	GCTGCCAGCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTGCCAGCGTGTAAAACGACGGCCAGT	ION_M13-384D_GCTGCCAGCG_F01
spike_96D	G01	01G	1G	GCCAGTCCAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCAGTCCATTGTAAAACGACGGCCAGT	ION_M13-384D_GCCAGTCCAT_G01
spike_96D	H01	01H	1H	AACCGCACGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACCGCACGTTGTAAAACGACGGCCAGT	ION_M13-384D_AACCGCACGT_H01
spike_96D	A02	02A	2A	GTGCTCCGAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGCTCCGAGTGTAAAACGACGGCCAGT	ION_M13-384D_GTGCTCCGAG_A02
spike_96D	B02	02B	2B	TATCTCGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ATCTCGATTTGTAAAACGACGGCCAGT	ION_M13-384D_TATCTCGATT_B02
spike_96D	C02	02C	2C	ACGACATTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGACATTCTTGTAAAACGACGGCCAGT	ION_M13-384D_ACGACATTCT_C02
spike_96D	D02	02D	2D	TCTGCTTGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTGCTTGCCTGTAAAACGACGGCCAGT	ION_M13-384D_TCTGCTTGCC_D02
spike_96D	E02	02E	2E	CTAATACTTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TAATACTTATGTAAAACGACGGCCAGT	ION_M13-384D_CTAATACTTA_E02
spike_96D	F02	02F	2F	TAACGTTATC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AACGTTATCTGTAAAACGACGGCCAGT	ION_M13-384D_TAACGTTATC_F02
spike_96D	G02	02G	2G	AGTGTCGGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTGTCGGTTTGTAAAACGACGGCCAGT	ION_M13-384D_AGTGTCGGTT_G02
spike_96D	H02	02H	2H	TTACACCGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TACACCGTTTGTAAAACGACGGCCAGT	ION_M13-384D_TTACACCGTT_H02

spike_96D	A03	03A	3A	CAGCGAGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AGCGAGATTTGTAAAACGACGGCCAGT	ION_M13-384D_CAGCGAGATT_A03
spike_96D	B03	03B	3B	GATATTCGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATATTCGATTGTAAAACGACGGCCAGT	ION_M13-384D_GATATTCGAT_B03
spike_96D	C03	03C	3C	TCTGTGCAAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTGTGCAACTGTAAAACGACGGCCAGT	ION_M13-384D_TCTGTGCAAC_C03
spike_96D	D03	03D	3D	GCTGATATCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTGATATCCTGTAAAACGACGGCCAGT	ION_M13-384D_GCTGATATCC_D03
spike_96D	E03	03E	3E	TTCACATTAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TCACATTAGTGTAAAACGACGGCCAGT	ION_M13-384D_TTCACATTAG_E03
spike_96D	F03	03F	3F	TGGAATGTCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GGAATGTCATGTAAAACGACGGCCAGT	ION_M13-384D_TGGAATGTCA_F03
spike_96D	G03	03G	3G	GAGCCTAGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AGCCTAGCATGTAAAACGACGGCCAGT	ION_M13-384D_GAGCCTAGCA_G03
spike_96D	H03	03H	3H	TAATGAATAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AATGAATATTGTAAAACGACGGCCAGT	ION_M13-384D_TAATGAATAT_H03
spike_96D	A04	04A	4A	GCTCTCTCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CTCTCTCGTTGTAAAACGACGGCCAGT	ION_M13-384D_GCTCTCTCGT_A04
spike_96D	B04	04B	4B	GCGTGTTACA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CGTGTTACATGTAAAACGACGGCCAGT	ION_M13-384D_GCGTGTTACA_B04
spike_96D	C04	04C	4C	CCGAATTATG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGAATTATGTGTAAAACGACGGCCAGT	ION_M13-384D_CCGAATTATG_C04
spike_96D	D04	04D	4D	CCTAATCGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CTAATCGTTTGTAAAACGACGGCCAGT	ION_M13-384D_CCTAATCGTT_D04
spike_96D	E04	04E	4E	CTTAACCATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTAACCATTTGTAAAACGACGGCCAGT	ION_M13-384D_CTTAACCATT_E04
spike_96D	F04	04F	4F	TTGGAACAGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGGAACAGGTGTAAAACGACGGCCAGT	ION_M13-384D_TTGGAACAGG_F04
spike_96D	G04	04G	4G	ACAGCCAGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CAGCCAGTTTGTAAAACGACGGCCAGT	ION_M13-384D_ACAGCCAGTT_G04
spike_96D	H04	04H	4H	ATGTCGGCAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TGTCGGCAATGTAAAACGACGGCCAGT	ION_M13-384D_ATGTCGGCAA_H04
spike_96D	A05	05A	5A	TCTGTAGTAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTGTAGTACTGTAAAACGACGGCCAGT	ION_M13-384D_TCTGTAGTAC_A05

spike_96D	B05	05B	5B	CAGCCATTCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AGCCATTCTTGTAAAACGACGGCCAGT	ION_M13-384D_CAGCCATTCT_B05
spike_96D	C05	05C	5C	ACGGCACTAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGGCACTAATGTAAAACGACGGCCAGT	ION_M13-384D_ACGGCACTAA_C05
spike_96D	D05	05D	5D	AGACACGTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GACACGTGATGTAAAACGACGGCCAGT	ION_M13-384D_AGACACGTGA_D05
spike_96D	E05	05E	5E	CATATCTACG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ATATCTACGTGTAAAACGACGGCCAGT	ION_M13-384D_CATATCTACG_E05
spike_96D	F05	05F	5F	CACGACCATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ACGACCATTTGTAAAACGACGGCCAGT	ION_M13-384D_CACGACCATT_F05
spike_96D	G05	05G	5G	ATCCGAGCGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TCCGAGCGCTGTAAAACGACGGCCAGT	ION_M13-384D_ATCCGAGCGC_G05
spike_96D	H05	05H	5H	AGACTCTGCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GACTCTGCCTGTAAAACGACGGCCAGT	ION_M13-384D_AGACTCTGCC_H05
spike_96D	A06	06A	6A	TGTGATAGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GTGATAGCATGTAAAACGACGGCCAGT	ION_M13-384D_TGTGATAGCA_A06
spike_96D	B06	06B	6B	TAGGCCACGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGGCCACGTTGTAAAACGACGGCCAGT	ION_M13-384D_TAGGCCACGT_B06
spike_96D	C06	06C	6C	ACTGGACTAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTGGACTATTGTAAAACGACGGCCAGT	ION_M13-384D_ACTGGACTAT_C06
spike_96D	D06	06D	6D	TATCACCGTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ATCACCGTGTGTAAAACGACGGCCAGT	ION_M13-384D_TATCACCGTG_D06
spike_96D	E06	06E	6E	CCAATGATCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CAATGATCCTGTAAAACGACGGCCAGT	ION_M13-384D_CCAATGATCC_E06
spike_96D	F06	06F	6F	ACGAATATGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGAATATGATGTAAAACGACGGCCAGT	ION_M13-384D_ACGAATATGA_F06
spike_96D	G06	06G	6G	GACGTTCGAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACGTTCGAATGTAAAACGACGGCCAGT	ION_M13-384D_GACGTTCGAA_G06
spike_96D	H06	06H	6H	ACCGGCAAGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CCGGCAAGGTGTAAAACGACGGCCAGT	ION_M13-384D_ACCGGCAAGG_H06
spike_96D	A07	07A	7A	GATGCACTCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATGCACTCATGTAAAACGACGGCCAGT	ION_M13-384D_GATGCACTCA_A07
spike_96D	B07	07B	7B	ATTCGCGAGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TTCGCGAGCTGTAAAACGACGGCCAGT	ION_M13-384D_ATTCGCGAGC_B07

spike_96D	C07	07C	7C	TTAGCAACGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TAGCAACGGTGTAAAACGACGGCCAGT	ION_M13-384D_TTAGCAACGG_C07
spike_96D	D07	07D	7D	TAGTCGAGCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGTCGAGCTTGTAAAACGACGGCCAGT	ION_M13-384D_TAGTCGAGCT_D07
spike_96D	E07	07E	7E	ATCTCTTCGG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TCTCTTCGGTGTAAAACGACGGCCAGT	ION_M13-384D_ATCTCTTCGG_E07
spike_96D	F07	07F	7F	AGCCGCGTGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GCCGCGTGTTGTAAAACGACGGCCAGT	ION_M13-384D_AGCCGCGTGT_F07
spike_96D	G07	07G	7G	TGCAATTACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GCAATTACCTGTAAAACGACGGCCAGT	ION_M13-384D_TGCAATTACC_G07
spike_96D	H07	07H	7H	GAACTACATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG AACTACATTTGTAAAACGACGGCCAGT	ION_M13-384D_GAACTACATT_H07
spike_96D	A08	08A	8A	CCGATTAATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGATTAATTTGTAAAACGACGGCCAGT	ION_M13-384D_CCGATTAATT_A08
spike_96D	B08	08B	8B	CTAGCCAAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TAGCCAAGTTGTAAAACGACGGCCAGT	ION_M13-384D_CTAGCCAAGT_B08
spike_96D	C08	08C	8C	AACGATGTGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACGATGTGATGTAAAACGACGGCCAGT	ION_M13-384D_AACGATGTGA_C08
spike_96D	D08	08D	8D	TCTTATGATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CTTATGATTTGTAAAACGACGGCCAGT	ION_M13-384D_TCTTATGATT_D08
spike_96D	E08	08E	8E	TTGTGCCACC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT TGTGCCACCTGTAAAACGACGGCCAGT	ION_M13-384D_TTGTGCCACC_E08
spike_96D	F08	08F	8F	GTATACAAGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TATACAAGTTGTAAAACGACGGCCAGT	ION_M13-384D_GTATACAAGT_F08
spike_96D	G08	08G	8G	GTGTCATGAA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGTCATGAATGTAAAACGACGGCCAGT	ION_M13-384D_GTGTCATGAA_G08
spike_96D	H08	08H	8H	CTAATGTCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TAATGTCTTTGTAAAACGACGGCCAGT	ION_M13-384D_CTAATGTCTT_H08
spike_96D	A09	09A	9A	GTACTTGCCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TACTTGCCATGTAAAACGACGGCCAGT	ION_M13-384D_GTACTTGCCA_A09
spike_96D	B09	09B	9B	ACCACGTGAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CCACGTGAGTGTAAAACGACGGCCAGT	ION_M13-384D_ACCACGTGAG_B09
spike_96D	C09	09C	9C	TAGTTGGTCC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT AGTTGGTCCTGTAAAACGACGGCCAGT	ION_M13-384D_TAGTTGGTCC_C09

spike_96D	D09	09D	9D	CGACGGATCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GACGGATCTTGTAAAACGACGGCCAGT	ION_M13-384D_CGACGGATCT_D09
spike_96D	E09	09E	9E	ATAGAACGCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TAGAACGCGTGTAAAACGACGGCCAGT	ION_M13-384D_ATAGAACGCG_E09
spike_96D	F09	09F	9F	AATCTGATTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ATCTGATTGTGTAAAACGACGGCCAGT	ION_M13-384D_AATCTGATTG_F09
spike_96D	G09	09G	9G	CTTGTAATTG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TTGTAATTGTGTAAAACGACGGCCAGT	ION_M13-384D_CTTGTAATTG_G09
spike_96D	H09	09H	9H	AGACCTGTTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GACCTGTTCTGTAAAACGACGGCCAGT	ION_M13-384D_AGACCTGTTC_H09
spike_96D	A10	10A	10A	CGAGCGAAGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GAGCGAAGCTGTAAAACGACGGCCAGT	ION_M13-384D_CGAGCGAAGC_A10
spike_96D	B10	10B	10B	CGACAAGACT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC GACAAGACTTGTAAAACGACGGCCAGT	ION_M13-384D_CGACAAGACT_B10
spike_96D	C10	10C	10C	AACGGTTGAG	CCATCTCATCCCTGCGTGTCTCCGACTCAGA ACGGTTGAGTGTAAAACGACGGCCAGT	ION_M13-384D_AACGGTTGAG_C10
spike_96D	D10	10D	10D	GCCAAGGCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG CCAAGGCTCTGTAAAACGACGGCCAGT	ION_M13-384D_GCCAAGGCTC_D10
spike_96D	E10	10E	10E	GATCACACCT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATCACACCTTGTAAAACGACGGCCAGT	ION_M13-384D_GATCACACCT_E10
spike_96D	F10	10F	10F	GACGCCGAAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ACGCCGAATTGTAAAACGACGGCCAGT	ION_M13-384D_GACGCCGAAT_F10
spike_96D	G10	10G	10G	CAATACCTAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC AATACCTATTGTAAAACGACGGCCAGT	ION_M13-384D_CAATACCTAT_G10
spike_96D	H10	10H	10H	AGATCCGCTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GATCCGCTCTGTAAAACGACGGCCAGT	ION_M13-384D_AGATCCGCTC_H10
spike_96D	A11	11A	11A	CCGGCCTCTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGGCCTCTTTGTAAAACGACGGCCAGT	ION_M13-384D_CCGGCCTCTT_A11
spike_96D	B11	11B	11B	TACCTGAGGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT ACCTGAGGCTGTAAAACGACGGCCAGT	ION_M13-384D_TACCTGAGGC_B11
spike_96D	C11	11C	11C	GATGTCTTCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGG ATGTCTTCATGTAAAACGACGGCCAGT	ION_M13-384D_GATGTCTTCA_C11
spike_96D	D11	11D	11D	GGTCACGGAC	CCATCTCATCCCTGCGTGTCTCCGACTCAGG GTCACGGACTGTAAAACGACGGCCAGT	ION_M13-384D_GGTCACGGAC_D11

spike_96D	E11	11E	11E	GTGCAGCCGT	CCATCTCATCCCTGCGTGTCTCCGACTCAGG TGCAGCCGTTGTAAAACGACGGCCAGT	ION_M13-384D_GTGCAGCCGT_E11
spike_96D	F11	11F	11F	AGACAGAGCA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GACAGAGCATGTAAAACGACGGCCAGT	ION_M13-384D_AGACAGAGCA_F11
spike_96D	G11	11G	11G	ACGCTCATTA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CGCTCATTATGTAAAACGACGGCCAGT	ION_M13-384D_ACGCTCATTA_G11
spike_96D	H11	11H	11H	AGAGATAATT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GAGATAATTTGTAAAACGACGGCCAGT	ION_M13-384D_AGAGATAATT_H11
spike_96D	A12	12A	12A	TCACAGCGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGT CACAGCGATTGTAAAACGACGGCCAGT	ION_M13-384D_TCACAGCGAT_A12
spike_96D	B12	12B	12B	ACTTGCGGAT	CCATCTCATCCCTGCGTGTCTCCGACTCAGA CTTGCGGATTGTAAAACGACGGCCAGT	ION_M13-384D_ACTTGCGGAT_B12
spike_96D	C12	12C	12C	AGTTAGATTC	CCATCTCATCCCTGCGTGTCTCCGACTCAGA GTTAGATTCTGTAAAACGACGGCCAGT	ION_M13-384D_AGTTAGATTC_C12
spike_96D	D12	12D	12D	CTGGTGCGGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGC TGGTGCGGATGTAAAACGACGGCCAGT	ION_M13-384D_CTGGTGCGGA_D12
spike_96D	E12	12E	12E	TGACGCCTGC	CCATCTCATCCCTGCGTGTCTCCGACTCAGT GACGCCTGCTGTAAAACGACGGCCAGT	ION_M13-384D_TGACGCCTGC_E12
spike_96D	F12	12F	12F	ATTACTAAGA	CCATCTCATCCCTGCGTGTCTCCGACTCAGA TTACTAAGATGTAAAACGACGGCCAGT	ION_M13-384D_ATTACTAAGA_F12
spike_96D	G12	12G	12G	CATAATGGTT	CCATCTCATCCCTGCGTGTCTCCGACTCAGC ATAATGGTTTGTAAAACGACGGCCAGT	ION_M13-384D_CATAATGGTT_G12
spike_96D	H12	12H	12H	CCGTCACGCG	CCATCTCATCCCTGCGTGTCTCCGACTCAGC CGTCACGCGTGTAAAACGACGGCCAGT	ION_M13-384D_CCGTCACGCG_H12

PROTOCOL

Allele Specific Amplification using “Spiked” Genotyping-by-Sequencing

Overview

Genotyping-by-sequencing (GBS) is an approach for reduced representation sequencing of large and complex genomes. Using a restriction enzyme, a small portion of the genome can be reducibly captured and sequenced.

Often genetics research and molecular marker assisted selection in plant breeding has need for single marker assays rather than whole genome profiling.

Primer Design

The assay is designed as a nested PCR reaction that can be completed in a single reaction well. Each sample will have a unique barcode primer with M13(-21) tail sequence. A set of common primers for the target sequence are included that have the corresponding M13 tail on the forward primer and a tail for the reverse sequencing primer site on the reverse primer. The nested PCR reaction will produce fragments that a ready for sequencing. The sequencing read will first read through the barcode followed by the M13 sequence. The target SNP can be located directly after the forward target sequence primer or further down stream as long as it is within the read length of the sequencing platform.

Allele Specific Amplification

1. Normalize 5ul of DNA at 20 – 40 ng/ul in a 96 well plate
2. Add 4ul of M13 barcode primer (0.75 uM).
   Note: Each sample well will have a unique barcode primer.
3. Make Master Mix for whole plate volume
4. Add 8ul of PCR master mix to samples

Regent (Stock Concentration)	Reaction Volume (ul)	Full Plate Volume (ul) (x120)	Final Concentration
Buffer Stock (10x)	1.5	180	1x
$MgCl_2$ (50 mM)	0.75	90	2.5 mM
dNTP mix (2.5 mM)	1.2	1.2	200 uM (each)
Forward Tailed Primer (10.00 uM)	0.03	3.6	20 nM
Reverse Primer (10.00 uM)	0.3	36	200 nM
Taq polymerase (5.00U/ul)	0.1	12	0.33 U
H_20	3.62	434.4	
Master Mix Total	8	960	
DNA (20 to 40ng/ul)	5	-	100 – 200 ng
M13 Barcode Primer (0.75 uM)	4	-	200 nM
PCR reaction total volume	**15**	**-**	

PCR CONDITIONS

PCR	Based pm Annealing temperature – short	
1	95ºC - 5 min	
2	95ºC - 1 min	
3	57ºC - 20 sec	36 Cycles
4	72ºC - 40 sec	
6	72C, 10 min	
7	8C, forever	

Spiking of Amplicon library to GBS library

The target amplicon library should be added at a concentration of ~1% of the total GBS library.

1. Quantify GBS library using PicoGreen
2. Normalize GBS library to 50ul at 11 nM
3. Quantify amplicon library using PicoGreen
4. Normalize amplicon library to 1.1 nM
5. Add 5 ul of amplicon library to 50 ul of GBS library

Library	Volume	Conc.	Final Conc.
GBS	50 ul	11 nM	10 nM
Amplicon	5 ul	1.1 nM	0.1 nM
TOTAL	55 ul		10 nM

Appendix D - Supplementary Materials Chapter 5

In addition to Field Book, we have developed several other phenotyping apps (PhenoApps) that can be used for collecting, managing, and analyzing data.

1KK

1KK is an app designed to analyze seed lots. Its name comes from the one thousand (1K) kernel weight that is commonly used as a selection criterion in plant breeding programs. 1KK extracts seed morphology from images captured by phone and tablet cameras. A non-parametric algorithm is used to identify individual seeds for shape measurements (Figure D-1). Reference circles of known size are included on a background mat and translate the pixel measurements of seeds to actual size. Each individual seed length, width, and area is determined using the algorithm first implemented in SmartGrain (Tanabata et al., 2012). Data can be exported in a sample summary format or on a per-seed basis. For measurement of thousand kernel weight, the total number of seeds are counted and divided by the total weight. For weight measurements, the app is compatible with 1g resolution USB scales (Elane). With a properly-sized reference background, potato tubers and cassava roots can be imaged and measured (Figure D-2). 1KK is open source (https://github.com/trife/1KK) and available on the Google Play Store (https://play.google.com/store/apps/details?id=org.wheatgenetics.onekk).

Inventory

To assist with rapid inventory and weighing of seed stocks, we've developed an application to inventory and weigh barcoded seed samples. Inventory uses a USB Scale (Elane) to quickly weigh and categorize samples. In addition to the Box and Sample ID, a timestamp and the name of the inventory person are also collected. Data is exported to a text file that can be directly uploaded to a central database. Inventory is open source

(https://github.com/trife/Inventory) and available on the Google Play Store (https://play.google.com/store/apps/details?id=org.wheatgenetics.inventory).

Coordinate

Coordinate is a data collection app that is based on defining templates and then collecting data in grids created from those templates. Two templates included by default are for seed trays, used to organize planting samples, and DNA plates, used to associate a tissue ID with the well into which it is being collected. Templates can be created to include custom fields for grid metadata collection (e.g. Person, Date, etc.); the naming for rows and columns can be alphabetic or numeric; and rows, columns, or random cells can be excluded from data collection. All collected data is saved internally to the database and grids can be reloaded to continue collecting data or deleted if not needed. Coordinate is open source (https://github.com/trife/Coordinate) and available on the Google Play Store (https://play.google.com/store/apps/details?id=org.wheatgenetics.coordinate).

References

Tanabata, T., T. Shibaya, K. Hori, K. Ebana, and M. Yano. 2012. SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol. 160(4): 1871–80.

Figure D-1. A processed photo of a wheat seed lot. Seeds that are identified as being individual are outlined in red and morphological measurements are collected. Blue reference circles of known size are outlined in white and used to scale pixel measurements to empirical measurements.

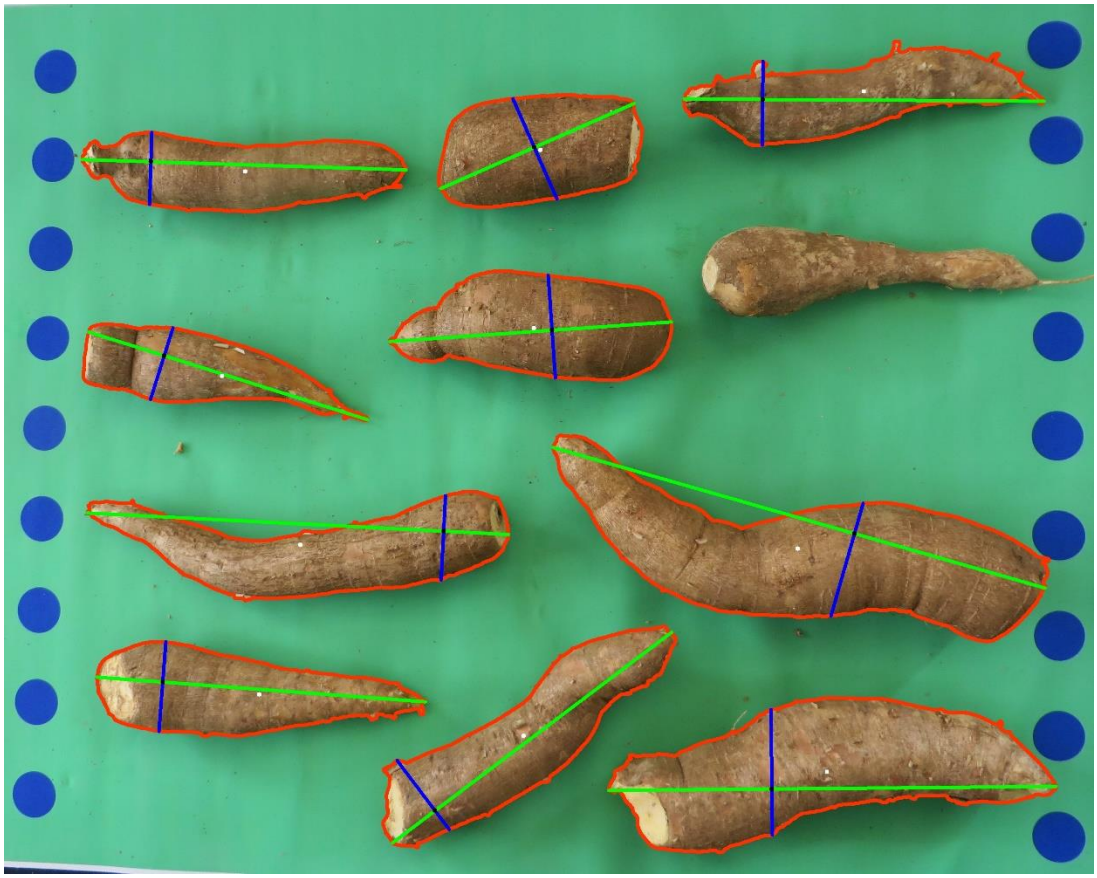

Figure D-2. A processed sample of cassava roots.