Temperate Corn Mid Density, Multiplexed SNP Panel.

Introduction:

Corn is one of the most economically important crops worldwide, serving as a staple food for human consumption, livestock feed, and as a raw material for various industries. It also serves as a model organism; the study of corn has been instrumental in the field of genetics, genomics, and plant physiology. Advances in corn genetics have facilitated the development of modern-day breeding methodologies, grain production practices, and genetic engineering**.**

Corn (maize, *Zea mays*) is an annual, monoecious grass in the family Poaceae. It is a diploid organism with a genome consisting of approximately 2.3 billion base pairs organized in ten chromosomes. The genome is characterized by a high proportion of repetitive DNA sequences, including transposable elements and regions with varying recombination rates.

The wild ancestor of corn, teosinte (*Zea mays spp. parviglumis*) is believed to have been domesticated in the Balsas River Valley of southwestern Mexico around 9,000 years ago (1, 2). Indigenous farmers recognized and enhanced desirable traits such as larger kernels, increased yield, and improved taste. Over time, improvements made by the early agricultural societies in Mesoamerica lead to the development of various landraces with distinct traits and adapted to different ecological niches and cultural preferences (**3**). The cultivation of maize gradually spread beyond its Mesoamerican center of origin and arrived in North America around 4,000 years ago. The Ancestral Puebloans, Mississippian cultures, and later, the Native American tribes adopted maize agriculture, shaping their societies and economies. The development of distinct landraces such as dent, flint, flour, and sweet corn further contributed to maize diversification (4). In South America, maize reached the Andes and the Amazon, where it became integrated into the agricultural practices of civilizations like the Inca, Moche, and Maya. The arrival of Europeans in the Americas in the 15th century facilitated the global dissemination of maize. Corn seeds were brought back to Europe, where they were initially cultivated as botanical curiosities. However, the crop's resilience, productivity, and nutritional value led to its wider adoption across Europe, becoming a vital food source during times of famine. Maize cultivation also spread to Africa, and Asia through European colonial expansion and trade routes. In Africa, maize become a staple crop in Kenya, Nigeria, and South Africa. In Asia, maize was integrated into traditional rice-based cropping systems. Today, maize holds immense cultural significance and is deeply embedded in the culinary traditions of various societies worldwide.

In the United States, approximately 3.5% of the corn grain harvested is used for cereals and other food products (5). This figure reflects a threefold increase in demand and consumption of maize-based food products since the 1970s, primarily because of the popularity of both Hispanic and gluten-free foodstuffs (6). Between the years 2015 to 2019, maize production in the United States totaled 363.2 billion kg on 33.5 million hectares with an average yield of 10,860 kg ha−1; this represents a 3.7-fold increase in per hectare corn productivity considering that the nationwide average yield was 2,950 kg ha−1 in 1956.

This increase reflects mostly advancements made in corn breeding methodologies, and improved genetics (7); Two important steppingstones count for much of this improvement, the first is the transition to hybrid corn. In the United States, maize germplasm is largely classified into three predominant heterotic groups: Stiff Stalk (SS), Non-Stiff Stalk (NS), and Iodent (IO) (8, 9, 10). Corn breeding and cultivation has evolved from open-pollinated varieties at the beginning of the last century to hybrid corn by exploiting hybrid-vigor (heterosis) between these complementary heterotic groups (11,12,13,14,15,16,17). Initially by implementing double-cross schemes and subsequently single-cross hybrids (18). Present day commercial maize is grown as a hybrid F1 cross of inbred lines from divergent heterotic groups and manifests a greater yield potential than their inbred parents.

The second steppingstone is the advent of marker assisted molecular breeding methodologies. Molecular markers have evolved over the past 80 years through sampling and comparing genomes. Over time, technology for interrogating genetic variation has progressed, and many DNA molecular marker systems were developed. Consequently, so did the resolution of the genomic picture the markers can depict. Single Nucleotide Polymorphisms (SNPs) have emerged as the ultimate molecular marker. SNPs are single nucleotide changes that are heritable, codominant, and distributed with relatively high frequency throughout eukaryotic genomes. SNPs can be the causative mutation that directly affects a phenotype or can be associated with such mutations. Marker-assisted selection (MAS) technologies offer an effective and accelerated approach for identifying and selecting desired traits in corn breeding programs by:

- allowing the identification of desired traits at early developmental stages, reducing the time and resources required for phenotypic evaluation.
- Precision and Efficiency: MAS enables the selection of individuals with the desired alleles with higher accuracy and efficiency.
- Increased Genetic Gain: MAS facilitates the stacking of multiple favorable alleles, leading to faster genetic gain in breeding programs.
- Enhanced Selection Intensity: MAS provides the opportunity to select individuals with favorable traits that may not be apparent through phenotypic evaluation alone.

We describe here in a mid-density SNP panel, designed to be used for marker assisted breeding applications of Temperate corn and its derivatives.

The SNP Panel:

The Temperate Corn mid density SNP panel is made of 2490 markers. The SNPs span the 10 maize chromosomes with an average of 249 SNPs per chromosome and a range of 178 SNPs on chromosome 10; the smallest chromosome, and 381 on the largest chromosome; chrome 1. The distance between adjacent SNPs ranges from 0.67 Mbp to 1.13 Mbp and averages 0.88 Mbp (Table 1). See Appendix A for a complete listing of the markers and their positions.

SNP selection has been done by Prof. Shawn Kaeppler's group at University of Wisconsin- Madison. SNP selection was carried out with the objective of identifying a set SNPs that segregate in dent maize, are

potentially associated with important agronomic traits, and are spaced evenly throughout the maize genome. Markers were selected based on both exome capture and genotype-by-sequencing markers previously generated for a six-parent synthetic MAGIC population (19, 20), and a set of 12 biparental populations formed from a factorial cross of seven parents. All populations are derived from inbreds adapted to temperate environments and therefore, all the SNPs in the panel were observed segregating in North American Dent germplasm.

Initial SNP identification was done using ridge regression (21) and obtaining the top 200 genetic markers with the largest effect for various traits of agronomic importance measured in both hybrids and inbreds. The regression results pointed to multiple SNPs in several genomic regions to be associated with complex traits due to linkage disequilibrium. To select one marker within each of these regions of importance and ensure the SNPs are evenly spaced, we performed step wise regression on the markers per combination of phenotype and population. Using this method, approximately 30 to 40 markers on average are identified per phenotype. Genetic mapping has pointed to additional SNPs associated with agronomic traits which completed the SNP list.

Marker-trait association analyses were performed across Stiff Stalk and non-Stiff Stalk mapping populations. However, none of the regions included in the final SNP selection were validated and further testing is required to confirm the effect of each SNP on agronomic trait of interest, therefore, trait associations are not listed on Appendix A until further validation. Markers in the panel are predicted to be associated with Agronomic traits such as plant growth and development, morphological features of maize kernels, cobs, and ears, flowering time, plant height and ear height, ear length and width, cob length and width, kernel mass, and disease resistance. A small number of SNPs were also associated with grain yield and grain moisture. Marker-trait association analyses for morphological traits of maize kernels, cobs, and ears were only performed among inbred lines while flowering time and plant and ear height association tests were conducted among both hybrids and inbred lines. Grain moisture and grain yield association tests were only conducted in hybrids when inbred lines were crossed to a tester in the opposite heterotic group.

A final subset of SNPs was included that are evenly spaced throughout the 10 maize chromosomes and were highly polymorphic across a diverse set of Stiff Stalk lines.

The panel was validated experimentally using germplasm provided by Prof. Shawn Kaeppler's laboratory at University of Wisconsin- Madison (188 lines), and a collection of lines put together by the Organic Research and Extension Initiative (OREI), coordinated by Prof. Paul Scott of Iowa State University and the USDA-ARS (94 lines). The University of Wisconsin samples were assembled to represent the three main corn heterotic groups, including 17 Iodent lines, 30 nonstiff stalk lines, and 41 Stiff stalk lines. In addition, a group of 56 Wisconsin Breeding Lines and 42 publicly released lines from the Genome Enhancement of Maize (22) project were included. Altogether, the participating lines can also be broken into 98 public lines and 90 ex-PVP lines.

The OREI maize varieties have traits suitable for organic production methods, for example, native resistance to insects and diseases since transgenes cannot be used, genes that confer cross incompatibility, to reduce possible GMO contamination by way of outcross pollination, and nutritional

quality as organic corn is used for food or feed. The lines are mostly derived from Iowa stiff stalk synthetic and Suwan-1 tropical population (23) that was developed in Thailand and adapted to the US by Arnel Hallauer.

Overall, the results of the validation experiment, analyzing the 282 samples with the 2490 SNP panel exhibited a call rate of 90%. The genotyping results indicate that the average Minor Allele Frequency of the markers does not differ significantly between the OREI and the Univ. of Wisconsin groups of samples and is averaging 0.3±0.22 amongst the samples. Taken as a whole, an 88.6% polymorphism was observed. A comparison of the number of polymorphic markers between any two lines was conducted as a measurement of the informativeness of the panel, and its potential usefulness for temperate corn breeding (Figure 1). The average number of polymorphic markers in across all pairwise comparisons was 1403 SNPs, which corresponds to 56.35% polymorphism. More than 70% of all pairwise comparisons had at least 1500 polymorphic markers.

Chromosome	N/Chromosome	Average Distance between Adjacent SNPs (Mbp)
1	381	0.8065
2	347	0.6914
3	256	0.8702
4	225	1.0725
5	324	0.6986
6	161	1.0822
7	213	0.8666
8	271	0.6707
9	178	0.8603
10	134	1.1355
Total:	2490	
Average:	**249**	**0.8755**

Table 1. Number of markers per chromosome and the average distance in Mbp between adjacent SNPs. The 2490 SNPs are evenly distributed among the 10 corn chromosomes with an average of 249 markers per chromosome.

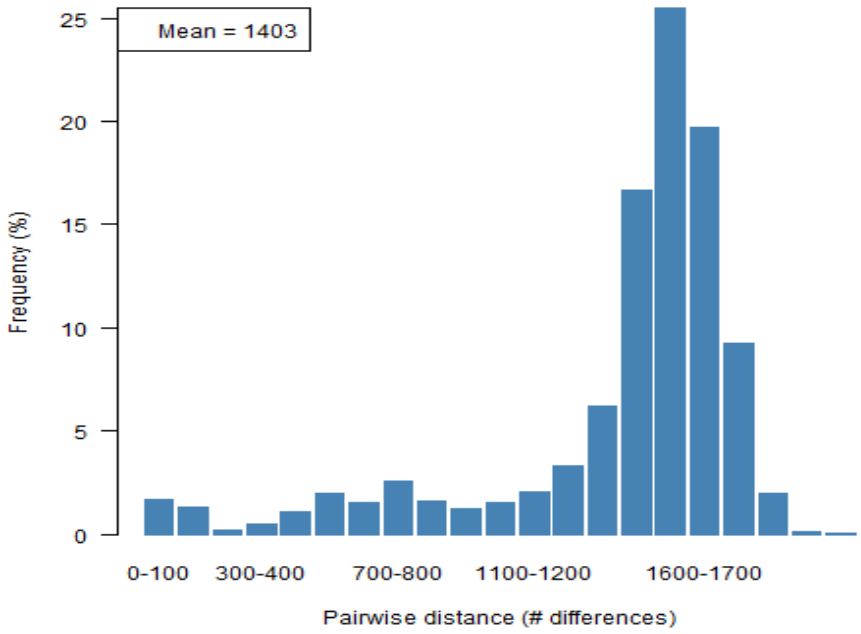

Figure 1: Frequency distribution of the number of polymorphic markers in pairwise comparisons amongst every two of the 282 lines included in the validation experiment. On average there are 1403 polymorphic markers between any pair of lines, or an average of 56.35% polymorphism.

PlexSeq™: The mid-density, multiplexed, SNP genotyping platform.

Several attributes of the PlexSeq™ process contribute to its unique value as a genotyping platform:

- The proprietary multiplexing algorithm, PlexForm™: The software designs all possible primers around all requested SNPs. Artificial intelligence algorithms identify the optimal sets of primers that can be mixed in one PCR amplification reaction.

- Once the amplifications are completed, the amplicon mixture is equivalent to barcoded libraries produced from other NGS methods. The process is unique because the samples produce amplicon libraries that are equivalent in concentration and do not require any additional equalization steps. A mixture of all the libraries is subjected to one bead cleanup and are loaded onto the sequencer. The process saves time, plasticware, and expenses.

- The method requires only minute quantities of crude DNA that can be isolated from a variety of tissues, enabling a quick and inexpensive DNA isolation process to start the genotyping workflow.

The PlexSeq™ workflow consists of:

- Crude DNA isolation
- Primary PCR: highly multiplexed, low volume (3ul) PCR amplifications
- Secondary, barcoding PCR amplifications
- Pooling: barcoded amplicons are combined into one tube, purified, and quantitated
- Sequencing on an NGS sequencer
- This relatively simple workflow is amenable to automation; all steps can be carried out on liquid handlers and high-capacity thermocyclers. This enables high throughput genotyping.

Once the sequencing is complete, a proprietary allele frequency-based genotype calling analysis software, PlexCall™, provides an automated sequencer to data workflow. This java-based software is tuned for each assay and is fully automated based on only the sequencing output files and a sample sheet indicating sample location on the plate.

Two other features make PlexSeq™ a unique fit for molecular breeding and seed QA. These applications typically require the genotyping of a large number of individuals. AgriPlex Genomics' extensive collection of unique barcode combinations allows for the simultaneous sequencing of thousands of individuals, limited only by the sequencer's capacity. Similarly, molecular breeding may require the addition or substitution of only some of the SNP markers as the individual breeding program advances or among programs as the parental-lines genetic makeup evolves. Those changes in SNP composition are also required for Quality Assurance (QA) applications as the diversity of the germplasm changes. The panel, being a collection of PCR primers not tethered to a surface (e.g.: chips) provides the flexibility to dynamically customize and alter the composition of the markers in the panel so it best fits the germplasm or the application. The Temperate Corn Panel is available as a service from AgriPlex Genomics and is also available as a kit to be used by in-house genotyping laboratories.

Applications

Genomic selection:

The Temperate Corn SNP panel will enable genomic selection; the level of polymorphism observed indicates that there is a high likelihood of obtaining sufficient polymorphic markers for imputation to a higher marker density level. The average Genomic Selection predictive ability will vary for different combinations of parental lines and in different years.

The combination of rapid, cost-effective genotyping of a prediction population during the last generation of line fixation saves expenses on the cost of field space for seed increase and allows rapid recycling of progeny as parents.

Marker-assisted backcrossing:

The Temperate Corn SNP panel can be used for background recovery estimates in marker-assisted backcrossing programs. The combination of informative background markers and a selection of trait

markers allows for the accurate estimation of background recovery, ensures recovery of valuable genes from the recipient line, and can provide additional confirmation that a target gene is carried by the selected progeny. Background selection can reduce by 2 or more the number of backcross generations required to achieve >95% recipient parent recovery.

QTL mapping:

While not the primary target application, the Temperate Corn SNP panel can be used for biparental mapping purposes. The density of polymorphic markers may, in some cases, be lower than what is desirable (largely dependent on the parents involved), however, the panel does provide an option for genotyping much of the genome. Any remaining gaps could then be filled in with other marker systems or by further panel customization.

Seed purity and hybridity:

The genome coverage of the Temperate Corn SNP panel includes the number of markers that will allow for the identification of diagnostic marker subsets for seed quality applications such as genetic purity testing (uniformity) and varietal identification in commercial production operations.

Conclusions

The Temprate Corn SNP panel is providing an excellent, cost-effective solution for applications requiring mid-density SNP numbers over any number of sample throughput. The panel fits with rapid line fixation protocols due to its low cost-per-sample and rapid turnaround time. This panel will allow for major-locus selection as part of genomic selection, or backcross introgressions amongst maize lines, or the transfer of genes and QTLs of interest from exotic germplasm into elite lines. The panel presents a valuable tool for saving critical time, expenses, and shortening the “time to market”. The panel primarily enables molecular breeding applications; however, the genome coverage expands its usefulness in a range of other research applications.

The flexibility of the PlexSeq platform permits continual revision and upgrading of the panel, ensuring the process keeps pace with the current breeding and seed production needs of the Maize breeding and research communities.

References:

1. Doebley, J. F. (2004). The genetics of maize evolution. Annual Review of Genetics, 38, 37-59.
2. Matsuoka, Y., et al. (2002). A single domestication for maize is shown by multilocus microsatellite genotyping. Proceedings of the National Academy of Sciences, 99(9), 6080-6084.
3. Smith, B. D. (1997). The initial domestication of maize in eastern North America. In The Origins of Agriculture in the Lowland Neotropics (pp. 111-130). Academic Press.
4. Wilkes, H. G. (2004). The origin and spread of maize in the New World. In The Cambridge World History of Food (pp. 184-197). Cambridge University Press.

5. USDA–National Agricultural Statistical Service, 2019. https://downloads.usda.library.cornell.edu/usda-esmis/files/k3569432s/sj139j59z/1257b842j/cropan20.pdf
6. Seth A. Tolley, Amritpal Singh and Mitchell R. Tuinstra, Heterotic Patterns of Temperate and Tropical Maize by Ear Photometry. Front. Plant Sci., 14 June 2021, Sec. Plant Breeding, Volume 12 - 2021 | https://doi.org/10.3389/fpls.2021.616975
7. Duvick, D. (2005). The contribution of breeding to yield advances in maize (Zea mays L.). Adv. Agron. 86, 83–145. doi: 10.1016/S0065-2113(05)8 6002-X
8. Nelson, P., Coles, N., Holland, J., Bubeck, D., Smith, S., and Goodman, M. (2008). Molecular characterization of maize inbreds with expired U.S. plant variety protection. Crop Sci. 48, 1673–1685. doi: 10.2135/cropsci2008.02.0092
9. Beckett, T., Morales, A., Koehler, K., and Rocheford, T. (2017). Genetic relatedness of previously Plant-Variety-Protected commercial maize inbreds. Plos One 12:e0189277. doi: 10.1371/journal.pone.0189277
10. White, M., Mikel, M., de Leon, N., and Kaeppler, S. (2020). Diversity and heterotic patterns in North American proprietary dent maize germplasm. Crop Sci. 60, 100–114. doi: 10.1002/csc2.20050
11. Shull, G. (1908). The composition of a field of maize. J. Hered. 4, 296–301. doi: 10.1093/jhered/os-4.1.296
12. Shull, G. (1909a). A pure-line method in corn breeding. J. Hered. 5, 51–59. doi: 10.1093/jhered/os-5.1.51
13. Shull, G. (1909b). Hybridization methods in corn breeding. J. Hered. 1, 63–72. doi: 10.1093/jhered/1.2.98
14. Shull, G. H. (1911). The genotypes of maize. Am. Soc. Nat. 45, 234–252.
15. Shull, G. H. (1914). Duplicate genes for capsule-form in Bursa bursa-pastoris. Z. Indukt. Abstamm. Vererbungsl. 12, 97–149. doi: 10.1007/BF01837282
16. East, E. (1909). The distinction between development and heredity in inbreeding. Am. Nat. 43, 173–181. doi: 10.1086/279042
17. East, E. (1936). Heterosis. Genetics 21, 375–397
18. United States Department of Agriculture, National Agricultural Statistical Service (USDA-NASS), https://www.nass.usda.gov/
19. Arrones A, Vilanova S., Plazas M., Mangino G., Pascual L., José Díez M., Prohens J. and Gramazio P. (2020). The Dawn of the Age of Multi-Parent MAGIC Populations in Plant Breeding: Novel Powerful Next-Generation Resources for Genetic Analysis and Selection of Recombinant Elite Material. Biology, 9, 229; doi:10.3390/biology9080229
20. Michel K. J., Lima D. L., Hundley H., Singan V., Yoshinaga Y., Daum C., Barry K., Broman K. W., Buell C. R., de Leon N., Kaeppler S. M. (2022) Genetic mapping and prediction of flowering time and plant height in a maize Stiff Stalk MAGIC population, GENETICS, 221(2), iyac063 https://doi.org/10.1093/genetics/iyac063
21. Hoerl, A. E. (1962). "Application of Ridge Analysis to Regression Problems". Chemical Engineering Progress. 58 (3): 54–59.
22. Genome Enhancement of Maize (GEM) project, https://www.ars.usda.gov/midwest-area/ames/plant-introduction-research/home/germplasm-enhancement-of-maize/

23. Wu X., Wang A., Guo X., Liu P., Zhu Y., Li X., and Chen Z. (2019) Genetic characterization of maize germplasm derived from Suwan population and temperate resources. Hereditas 156:2 https://doi.org/10.1186/s41065-018-0077-1