Spring 2020

Testcross Seed Optimization: Case Study of Transgenic Maize

Komlan Avochinou

Testcross Seed Optimization: Case Study of Transgenic Maize

By

Komlan S. Avochinou

A creative component submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Plant Breeding

Program of Study Committee:

Anthony Assibi Mahama

Thomas E. Lubberstedt

Iowa State University

Ames, Iowa

2020

TABLE OF CONTENTS

Nomenclature (abbreviations used)

GMO: Genetically Modified Organism

Event: DNA sequence incorporated into the target genome and the specific point(s) of insertion.

Construct: an artificially constructed segment of nucleic acid which is going to be transplanted' into a target tissue or cell

USDA: US Department of Agriculture

TSP: Testcross Seed Production

MCDM: Multiple Criteria Decision Making

HC: Hand crossing

ICB: Isolated Crossing Block

ACKNOWLEDGMENT

I want to thank the committee chair, Dr. Anthony Assibi Mahama, and committee member Dr. Thomas Lubberstedt for their guidance and support throughout this research. Also, I would like to thank especially Michelle Zander, for all the help and guidance through this program and classwork. Finally, thanks to my wife, Ghislaine, for her patience, understanding, and support throughout the program.

ABSTRACT

Efficient resource allocation is a crucial component of every successful breeding program. Especially in transgenic maize production, the mandatory regulatory requirements in the industry to bring these genes to the market add additional costs that quickly raise the overall research budget. Twelve years of nursery data of a transgenic startup firm were analyzed, and a production model developed using multiple criteria decision making (MCDM) approach to optimize production inputs and outputs. The total number of events trialed nearly doubled without additional land requirement due to savings resulting from a reduction in testcross seed production.

INTRODUCTION

Genetically modified organisms (GMOs) are globally grown in larger quantities than ever before around the world. Unlike a conventional breeding program, the transformed lines require additional guided field-testing through the evaluation of testcross progenies before any potential validation (Kok et al., 2018).

Testcross seed production (TSP) is expensive, yet it is a vital step in the inbred line or event selection process in transgenic maize (Zea *mays* L.) research. TSP is required for replicated trials over targeted testing environments to validate the performance of a line, and especially in the case of transgenic maize, the effect of the gene under study (Fernandez-Cornejo & Caswell, 2006). In transgenic maize development, lines bearing the gene of interest are crossed to a homozygous common tester, and the resulting hybrids are subjected to more stringent testing based on government regulations, from gene discovery through field evaluation (Mumm, 2013). The combination of rigid regulatory requirements coupled with rigorous guidelines for field evaluation (Jonas, 2015) adds substantial cost to the research budget before any potential elite event even reaches the commercial market (Prado et al., 2014).

While the costs of gene sequencing become more and more affordable (Schatz et al., 2012), the costs for field testing have significantly increased over time. Much of the latter costs are caused by the increase in land requirement as the breeding program grows, and more events become available for field testing. According to the 2018 USDA survey, cropland value has a sustained overall positive trend (USDA, 2018). That is, as more events come down the transgenic pipeline, so there is pressure to increase the TSP for field trials, which can quickly result in a trial bottleneck in case of limited budgets.

Furthermore, in transgenic line development, the transformability of a line is significantly genotype-dependent, and very few genotypes are reported to be easily transformable (Rasha et al., 2013). The selection of a transformable genotype in a breeding program is generally not depending on superior agronomic attributes but instead on amenability for tissue culture and transformability (Yadava et al., 2017). Consequently, the transformed inbred lines bearing the gene of interest may lack key agronomic attributes such as high yield potential or disease resistance, that qualitatively and quantitatively affect the resulting field trials for event validation. In addition to these potential genetic disadvantages, non-transgenic testcross seeds must be produced alongside to serve as a check in the experimental design to better evaluate the performance of the trait under study (Kok et al., 2018). While this latter part is done conventionally, the volume of these check seeds produced must be relative not only to the size of the trial but the chosen experimental design. Numerous literature and agricultural extension manuals discussed and emphasized strategies on the best management practices to maximize seed return in maize hybrid seed production (Culy et al., 1991; Macrobert et al., 2014). The seed return maximization, while crucial at every step of the TSP, should not be the ultimate goal, regardless of the breeding stage or the research goal.

Ideally, within a breeding program, the nursery department, and the trial department must coordinate to ensure that the product development goals are approached effectively in terms of testcross seed availability to run the trials, especially in the early phase of hybrid testing since untested materials are involved. However, having enough seeds to run a successful trial does not necessarily imply an effective or efficient level of TSP. The production of testcross seeds while meeting the seed availability goal, can often

far exceed a reasonable optimum TSP level; and the significant discrepancy between the needed testcross seeds to run the trial and the total quantity produced in the first place has failed full scrutinization because of lack of systematic approach to TSP.

The objective of this paper is to estimate the overlooked waste in testcross seed production in the early phase of inbred line development using a case study of the transgenic seed development program of a startup biotech firm. The paper also examines how a systematic allocation of resources through a model tailored to a well-defined research goal can significantly minimize production inputs while achieving the initial experimental trial target, and so improve the research budget through shifting of resources.

MATERIALS AND METHODS

Field Data

The field data used to develop the improvement model were based on winter nursery archive records compiled from a single location, Santa Isabel, Puerto Rico from 2005 through 2017, and the yield trial data collected at three locations in Iowa, Illinois, and Nebraska.

Nursery layout and field maintenance

From 2005 to 2013, the transgenic planting pattern consisted of a male row (M1) alternating with two pairs of identical female rows (F1, F1, F2, F2) in a hand crossing block (HCB). The male row (M1), the transgenic line was delayed by 100 units GDD (Growing Degree Days) relative to the female rows at planting to allow better nicking (Fig 1a). The nursery plots were 6 m long with 0.76 m plot-to-plot alleyway and comprised of 30 plants

with 0.15 m within row spacing. The inbred line A188 bears the gene of interest, is highly transformable, but lacks critical agronomic attributes (D'Halluin et al., 1992). A188 was used as a male and RB01, a conventional line, as a female. Twenty percent of the total nursery was devoted to the production of near-isogenic hybrid lines by crossing the non-transformed male A188 to the conventional female. Each planting season, the field was irrigated to stabilize the performance of the male A188 as a weak line that is susceptible to disease and drought. The nursery was sprayed with preemergence herbicide Prowl at 3.55 l/ha [use SI units: ha instead of acres] and two applications of insecticide Coragen and Mustang Maxx, respectively, at 0.52l/ha and 113.6 ml/ acre. An N-P-K (10-34-0) fertilizer was applied at 226.75 kg/ha at planting, and a minimum isolation distance of 183 m was established around the nursery in compliance with USDA-APHIS regulation on GMO crop development.

From 2014 through 2017, the same field maintenance was performed, and same biometric data collected on the plots with the exception that, (a) the row pattern was changed from two pairs of different event rows (Fig 1a) to four rows of different female (F1, F2, F3, F4) (Fig 1b), (b) the male rows were doubled, and the second-row delay planted when the first male row emerged to increase pollen availability, (c) the transformed line A188 was used as a female instead, (d) the total conventional crossing block (check seed) produced revised from 20% of the nursery to a conditional size using Pierre Dagnelie check guidance reference formula (Dagnelie, 2012), $n_0 = n\sqrt{p}$ where n_0 represents the number of repetition of the check, n the number of repetition of the events selected for trials for a total of p events. All female rows were detasselled before pollen shed to avoid self-pollination, and the male rows removed after pollination.

Biometric data

At maturity, each female row was hand-harvested, individually shelled, and passed through color sorter VMEX™ MTX-160-33 that systematically ejected diseased and cracked kernels based on established calibrations to improve seed quality. The final usable hybrid seeds were counted using the Marvin™ Digital Seed Analyzer.

The Marvin seed analyzer was connected to a precision scale that estimates the total weight of the main bag based on 1000 kernels weight generated from seed sample spread over the analyzer platform.

While the seed analyzer output provided data on many seed characteristics or traits such as average grain length, surface area, and width of the seed, the primary traits of interest for the model development are (1) average kernel weight for each event, (2) the 1000 kernel weight, and (3) total usable seed quantity available for trial.

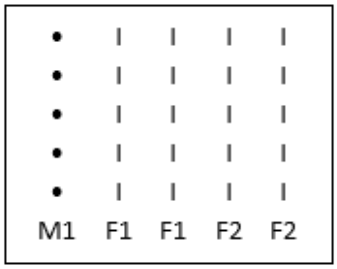

Fig 1a

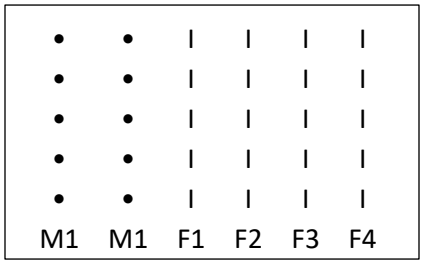

Fig 1b

Fig 1a and b: Planting arrangement used in testcross seed production. The dotted lines represent the male rows, and the dashed lines represent the female rows.

The scope of the field trial

The preliminary trial was evaluated in simple lattice using a two-row plot of 6.1 m length, 0.76 m spacing between rows, at three locations, and three replications per

location. Each planting packet was filled with 90 kernels with the expectation of 36 kernels planted per row on average, using SRES™ vacuum research planter with splitter cone that equally splits each packet of 90 kernels into two rows.

Data analysis and model simulation

Data were analyzed using R programming, with factors considered significant at $\alpha = 0.05$. The model simulation took into account eight years of historical nursery data, and the estimate of optimal seed requirement (given the yield trial need), and the reduction in plot length was calculated using a combination of the following three equations:

t = loc x rep x k (theoretical need) (I)

$h \times p \geq t$ (II)

$p = \left(\frac{l}{s} + 1\right)$ (III)

$t' = t(2 + B - A)$ (realistic need) (IV).

Equation (I) defines the theoretical minimum testcross seed quantity t needed to justify an event entry into the trial, where *loc* is the number of locations selected for the trial, rep is the number of repetitions per line, and k, the maximum testcross seed required for an experimental unit (two-row plots). In equation (II) t is reformulated as a product of h and p, where h is the average number of seed produced per cross and p the number of plants per row.

This production target value (t) being environment-dependent, along with deviational factors such as the germination percentage A of the seed after harvest and the average rejection percentage B after seed conditioning, was considered in the estimate of the realistic seed requirement per event (t') in equation (IV).

Equation (III) indicates the intrinsic relation between the maximum number of plants *p* per row, and it is a function of the plot length *l* and plant to plant spacing *s*.

RESULTS AND DISCUSSION

Between 2005 and 2013, the analysis of pollinated ear per row data showed high variability within a female row, and the variability increased as the row was positioned further away from the male row in the hand-crossing design (Fig 2). While the average pollinated ear counts per row decreased from F1 to F4, the variability increased from F1 toward F4 (Fig 3).

Despite this high variability in the female rows as they are positioned further away from the male row, the pollinated ear per row count suggests testcross seed production that is, however, higher than the t value which is a theoretical minimum testcross seed needed per event to justify an entry into the trial (Equation I). Similarly, R1, the first female row pollinated ears averaged 29 pollinated ears over 30 plants, an output three times higher than the t value. Furthermore, the near-isogenic non-transgenic counterpart testcross production as a check for the trial represented 20% of the total nursey and deviated significantly from the suggested ratio based on Dagnelie (2012).

From 2013 through 2017, in isolation crossing block design, the events per crossing block doubled, and gained a slight increase, while the total land allocated decreased little more than six times (Fig 4) because plot length was reduced from 6 m to 2.4 m.

The high within row variability and consistent pattern noted under the hand crossing design with a single male row suggests an early shortage of pollen sources to

cover all four female rows rather than lack of experienced pollinators. The row position effects within the seed maize production field are well-established (Culy et al., 1991). In commercial hybrid production, for example, the male and female inbred parents are carefully chosen based on not only their combining ability but other agronomic attributes such as pollen shed or per se yield performance of the female line that enhances the hybrid seed production. The disadvantage in the transgenic development programs, especially with startups and organizations with a limited budget, is the access to the highly transformable germplasm with good agronomic attributes.

Unlike the conventional breeding program where the most competitive and high performing hybrid is used as the check for the targeted zone, preliminary transgenic field trials suffer the disadvantage of using only their near-isogenic counterpart or null segregant as a check, the production of which must be done in a systematic way that takes into account the total events selected into trial, and the chosen replication number for the trial stage.

From 2013 to 2017, the choice of using the transgenic male as transgenic female while doubling the male row, increased the pollen source since the previous female is also a good pollen shedder. The first female row (F1) data from Table 1 suggests not only the potential of decent seed return as the pollen source is fully available but an overproduction of testcross seed to nearly eight times the t value which is the quantity of the testcross seed required per event to justify the event entry into the initial trial; in this particular case 810 kernels. Moreover, the simulation model (equation IV), taking into account the germination percentage A, and the rejection percentage B after quality

control, matches the measured data (Table 2) for a significant reduction in land used while maintaining the slight increase of transgenic rows.

Table 1: Statistics for Pollinated Row in female rows F1, F2, F3, F4, respectively, where F1 is the closest row to the male from left to right.

Statistic	F1	F2	F3	F4
Average seed/row	6480	6110	5052	3290
Average plant/row	29	26	21	14
SE_{mean}	0.01	0.04	0.07	0.13
CI_mean (0.95)	0.01	0.07	0.14	0.25
Variance	0.25	6.61	26.88	81.42
Std Dev	0.50	2.57	5.19	9.02
Coef. Var	0.02	0.10	0.24	0.60

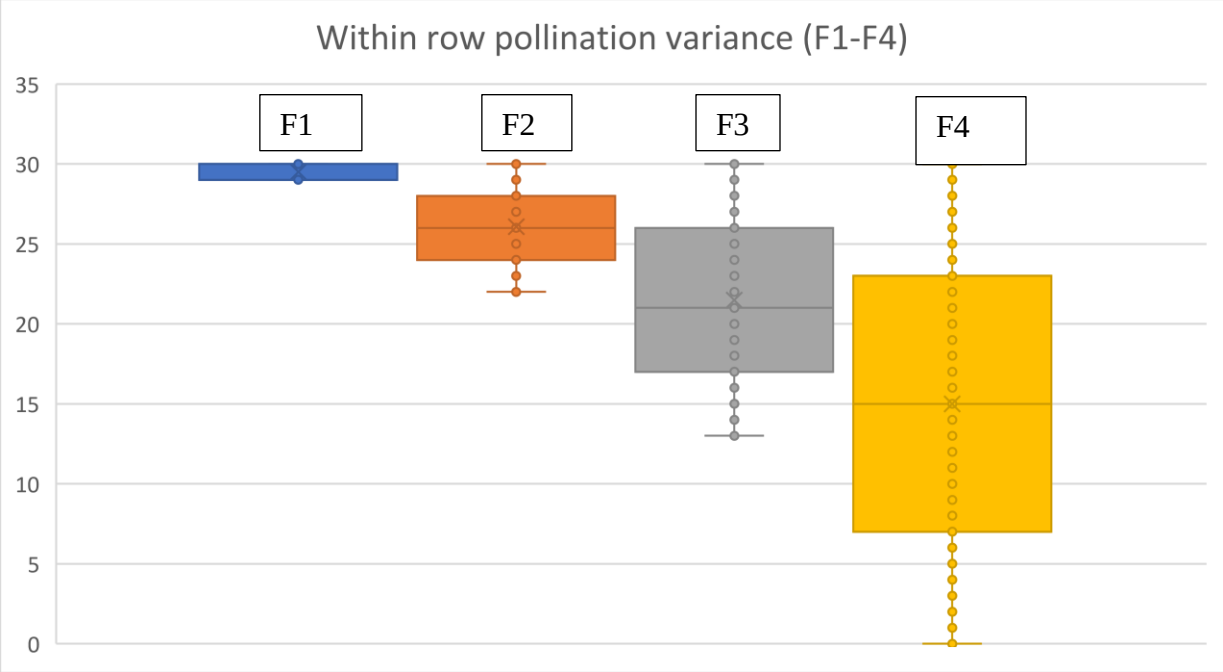

Figure 2: Pollination counts variance under hand crossing between 2005-2013

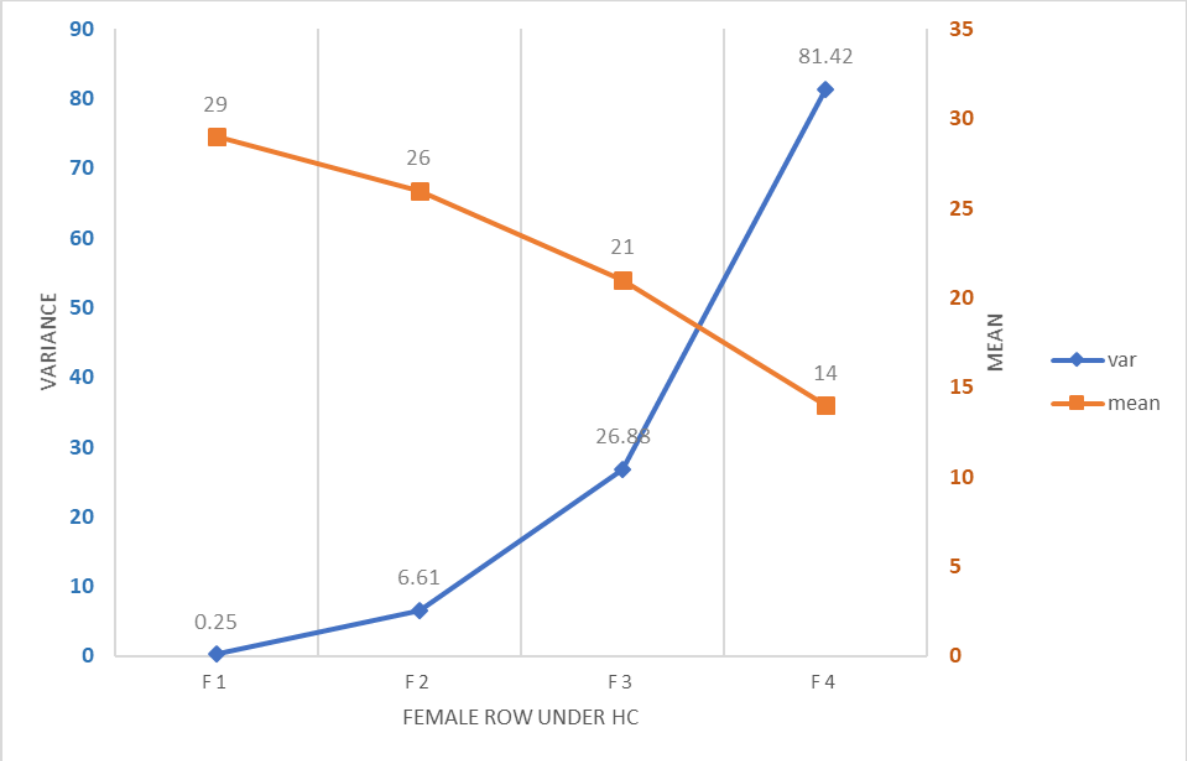

Figure 3: Mean and variance of pollination count on F1, F2, F3, F4 under HC

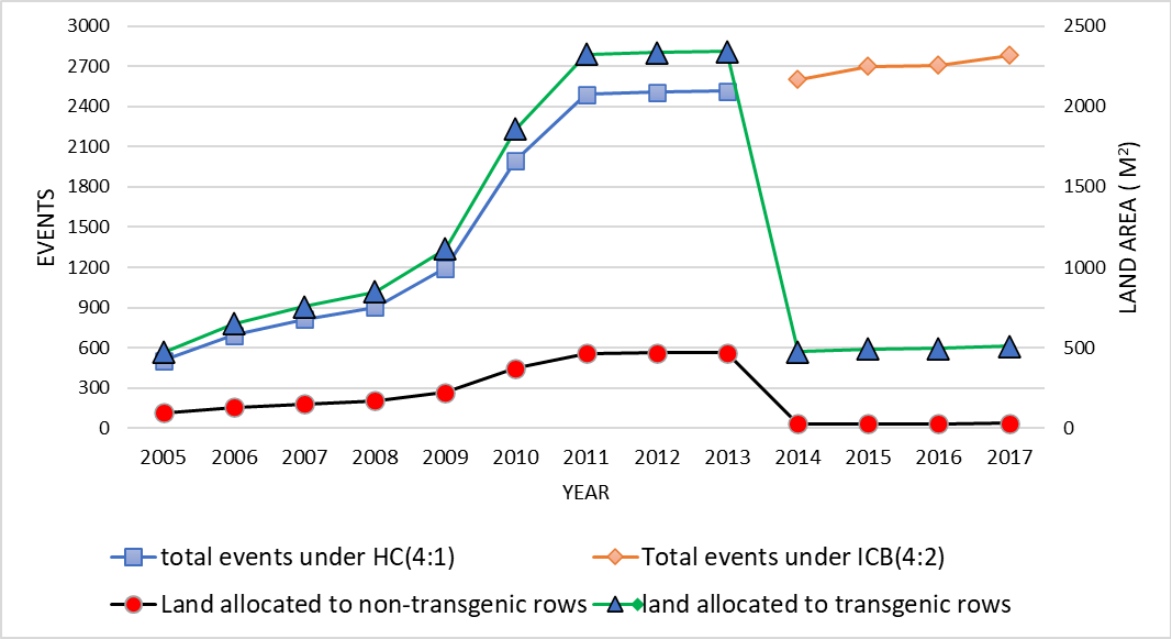

Figure 4: Testcross seed produced, and land allocated

Year	total events under HC(4:1)	Total events under ICB	Land allocated under HC (m^2)	Land allocated under ICB (m^2)	Land allocated to non-transgenic rows (m^2)	Land allocated to Isogenic rows under HC (m^2)	Land allocated to Isogenic rows under ICB (m^2)	land allocated to transgenic rows from HC to ICB (m^2)
2005	506		472		94	94		472
2006	692		651		130	130		651
2007	810		755		151	151		755
2008	904		848		170	170		848
2009	1192		1116		223	223		1116
2010	1998		1859		372	372		1859
2011	2492		2324		465	465		2324
2012	2506		2335		467	467		2335
2013	2514		2343		469	469		2343
2014		2602		476	28		28	476
2015		2698		493	28		28	493
2016		2706		495	29		29	495
2017		2780		508	29		29	508

Table 1: Summary of events fluctuation and land allocation from 2005 to 2013 under HC, and from 2014 to 2017 under ICB.

CONCLUSION

A combination of accurate inbred biometric data and environmental specificity was found to be the best predictor of testcross seed production level. While the per se performance of the male and female parent lines is proven not to be a good predictor of the hybrid performance, it is, however, a crucial factor in testcross seed optimization. The simulation model accurately reflects the benefit of tailoring testcross seed production to the stage of a transgenic breeding program considering the upfront cost related to transgenic crop regulation, namely gene discovery as cutting-edge technology, and field trial cost. Cost-effective nursery design, coupled with the choice of appropriate germplasm that is not only convenient when it comes to line transformability, but also good agronomic attributes, is mandatory.

As the world population growth increases and most of it is in the developing world, the transfer and adoption of the technology in these regions where the lack of funds to research is endemic, a built-in cost-saving procedure in the early stage of the research development will be a benefit.

REFERENCES

Culy, M. D., Edwards, C. R., & Cornelius, J. R. (1991). Row Position Effects within Seed Corn Production Fields on Yield and Quality of Inbred Corn. *Purdue Univ. Agric. E\p. Stn., and Pioneer Hi-Bred International. Journal Paper Published in J. Prod. Agric*, *4*(3), 373–376.

Dagnelie, P. (2012). Principes d'expérimentation - Planification des expériences et analyse de leur résultats. In *Les presses agronomiques de Gembloux*.

Fernandez-cornejo, J., & Margriet Caswell. (2006). The First Decade of Genetically Engineered Crops in the United States. In *Economic Information Bulletin* (Vol. 11). Retrieved from www.ers.usda.gov

Jonas, P. (2015). *Compliance Management Systems*. 1–18.

Kok, E. J., Kleter, G. A., Visser, R. G. F., Slot, M. M., & van de Wiel, C. C. M. (2018). The assessment of field trials in GMO research around the world and their possible integration in field trials for variety registration. *Transgenic Research*, *27*(4), 321–329. https://doi.org/10.1007/s11248-018-0076-z

Macrobert, J. F., Setimela, P., Gethi, J., & Regasa, M. W. (2014). Maize Hybrid Seed Production Manual. In *Book*. https://doi.org/10.13140/RG.2.1.4791.6965

Mumm, R. H. (2013). A look at product development with genetically modified crops: Examples from maize. *Journal of Agricultural and Food Chemistry*, *61*(35), 8254–8259. https://doi.org/10.1021/jf400685y

Prado, J. R., Segers, G., Voelker, T., Carson, D., Dobert, R., Phillips, J., … Martino-Catt, S. (2014). Genetically engineered crops: from idea to product. *Annual Review of Plant Biology*, *65*, 769–790. https://doi.org/10.1146/annurev-arplant-050213-040039

Rasha, O. a., Matheka, J. M., Abdelbagi, A. M., & Machuka, J. (2013). Transformation of Tropical Maize with the NPK1 Gene for Drought Tolerance. *International Journal of Genetic Engineering*, *3*(2), 7–14. https://doi.org/10.5923/j.ijge.20130302.01

Schatz, M. C., Witkowski, J., & McCombie, W. R. (2012). Current challenges in de novo plant genome sequencing and assembly. *Genome Biology*, *13*(4), 243. https://doi.org/10.1186/gb4015

Yadava, P., Abhishek, A., Singh, R., Singh, I., Kaul, T., Pattanayak, A., & Agrawal, P. K. (2017). Advances in Maize Transformation Technologies and Development of Transgenic Maize. *Frontiers in Plant Science*, *7*(January), 1–12. https://doi.org/10.3389/fpls.2016.01949