Effects of a long-term recurrent selection program on the genetic structure of the BSSS maize population

by

Paulo Evaristo de Oliveira Guimarães

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Plant Breeding

Program of Study Committee:
Kendall R. Lamkey, Major Professor
Arnel R. Hallauer
E. Charles. Brummer
Fredric J. Janzen
Michael Lee

Iowa State University

Ames, Iowa

2001

Graduate College

Iowa State University

This is to certify that the Doctoral dissertation of

Paulo Evaristo de Oliveira Guimarães

has met the dissertation requirements of Iowa State University

Major Professor

For the Major Program

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

ABSTRACT v

GENERAL INTRODUCTION
- Introduction 1
- Dissertation Organization 3
- Literature Review 3
- References 16

CHANGE RFLP ALLELE AND GENOTYPIC FREQUENCIES AFTER HALF-SIB AND S_2 SELECTION IN THE IOWA STIFF STALK SYNTHETIC MAIZE POPULATION
- Abstract 21
- Introduction 22
- Materials and Methods 24
- Results and Discussion 27
- References 39

MOLECULAR MARKER DIVERSITY FOLLOWING HALF-SIB AND S2 SELECTION IN THE IOWA STIFF STALK SYNTHETIC MAIZE POPULATION
- Abstract 53
- Introduction 54
- Materials and Methods 56
- Results and Discussion 62
- References 72

GENERAL CONCLUSIONS
- General Discussion 92
- References 96

ACKNOWLEDGEMENTS

I dedicate this dissertation to my wife, Soraya, our sons, Gustavo and André, my father and mother, Arlindo e Madalena, and my mother in law (in memoriam), Irene. Their love, determination, encouragement, understanding, and patience were outstanding and essential to the completion of my studies and the happiness of our family.

It was an honor for me to have Dr. Kendall R. Lamkey as my major professor. His advice, friendship, understanding, and support during all of these four years were invaluable for the development of my study and research program. I wish to express great appreciation and gratitude to have Dr. Arnel R. Hallauer, Dr. Michael Lee, Dr. E. Charles Brummer, and Dr. Fredric J. Janzen as my graduate committee members. Their guidance, support, and suggestions were very useful for the development of my studies.

Special gratitude to my sponsor, EMBRAPA, the Brazilian Agency for Agricultural Research, for providing an outstanding support for my studies and also for my family expenses. Thanks also to Dr. Cleso A. P. Pacheco for his constant friendship and for taking care with special attention to all of my personal interests.

I enjoyed the great work and learning environment of the Corn Breeding Laboratory. I appreciate the sharing of experiences, encouragement, and friendship of Kebede Mulatu, Lori Hinze, Buppa Kongsamai, Dr. Joanne Labate, Chris Mack, Dr. Mathias Frisch, Dr. Milosav Babic, John Golden, Jim Sears, Dr. Jode Edwards, Dr. Goran Bekavac, Charles Foresman, Dr. Peter Guzman, Sandra Hagdorn, and Clinton Turnbull.

The cordial relationship with the Brazilian community was very important to my family and me. Special thanks for members of Swanson, Silva, Santos, Peternelli, Cheng, Michelini, Moreira, Scafi, Armenio, and Ferraz families for their friendship and encouragement.

ABSTRACT

Evaluating of Recurrent Selection (RS) programs can lead to increase knowledge of methods, populations, and traits and give support for better management of breeding programs. The objective herein was to evaluate the effects of seven cycles of half-sib selection followed by seven cycles of S_2 selection on the genetic structure of BSSS maize population. Individuals from BSSSP (progenitor lines), BS13(S)C0 (original S_2 selection), and BS13(S)C7 (7^{th} S_2 cycle) cycles were genotyped based on a sample of 105 RFLP loci.

Measures of genetic variation within (expected heterozygosity, number of alleles, average frequency of the most common allele, and proportion of polymorphic loci) and among (Principal Component Analysis and Nei's genetic distance, NGD) cycles of selection indicated: BSSSP has a considerable genetic variability, substantial loss of variation and increase of divergence over the cycles of selection, greatest loss of diversity occurred during the HS selection program, future cycles of RS are predicted to have narrow genetic variation, and low average effective population size was an important factor in loss of genetic variation.

Changes in allele frequencies for about 30% of the loci cannot be explained by genetic drift alone, suggesting that selection also was an important factor of variation. The majority of loci in C0 and C7 were in H-W equilibrium. Progenitor line Illinois Hy had a lower NGD to C0 and C7 and five of its unique alleles had frequencies significantly increased in later generations, indicating a selective advantage over the cycles of RS. Hybrid Hy x LE 23 showed the lowest NGD to C0 and C7 populations. NGD among parental lines was not a good predictor of single-crosses yield performance. A founder effect observed herein may explain partially reduced genetic gains during the S_2-selection period reported in other studies. Limited RFLP diversity in BS13(S)C7 suggests this population may not have enough

genetic variability to sustain significant long-term genetic gains per se for grain yield. RFLP data were useful tools to evaluate this RS program. However, much more information could be obtained about recurrent selection programs by integrating of molecular (a standard set of marker loci) and phenotypic data.

GENERAL INTRODUCTION

Introduction

A breeding program can be considered as a management process where the main goal is to use properly the genetic variation (resources) to develop better phenotypes (products). Efficiency in recurrent selection programs means continuous development of superior genotypes with minimal loss of genetic variation (Hallauer and Miranda, 1988). The genetic structure of a population is characterized by the allelic and genotypic frequencies of its gene pool. Selection, genetic drift, nonrandom mating, migration, and mutation can lead to alteration of genetic diversity within and among populations (Falconer and Mackay, 1996; Russell, 1996). Intense selection pressure, small effective population size (genetic drift), positive assortative mating, inbreeding, and complex interactions among them are the principal factors causing changes in the genetic structure of populations undergoing recurrent selection. Assessment of genetic variation in populations undergoing recurrent selection (RS) can provide information that results in a better understanding of populations, breeding methods, and traits undergoing selection. This knowledge should lead to better management of plant breeding programs.

Several maize populations have been improved by intrapopulation and reciprocal recurrent selection. Quantitative genetic analysis of phenotypic data has revealed vast amounts of genetic information for improved maize populations (see review in Hallauer and Miranda, 1988). Genetic components of variances, heritability, inbreeding levels, predicted and observed responses to selection, and combining ability are some of the measures used to estimate genetic diversity of populations undergoing selection. Isozymes and molecular markers, such as restriction fragment length polymorphisms (RFLPs), also have been used to

evaluate genetic variation in populations undergoing selection (Brown, 1971; Brown and Allard, 1971; Stuber and Moll, 1972; Stuber et al., 1980; Kahler, 1983; Revilla et al., 1997; Heredia-Diaz et al., 1996; Labate et al., 1997, 1999, and 2000; Popi et al., 2000). Polymorphism level, observed and expected heterozygosity, number of alleles, and genetic distances are examples of measures of genetic variation that these tools can provide (Nei, 1987).

Consistent with the genetic and environmental complexity of breeding programs, results based on quantitative genetic analysis of phenotypic data and also isozyme and molecular data are variable and dependent on diverse factors such as: population, breeding method, effective population size, number of cycles of selection, etc. Also, as indicated by Lynch (1995), genetic variation estimates based of phenotypic, isozyme, and molecular data are also subject to considerable sampling errors and the correlation between molecular and quantitative genetic estimates of genetic variation are expected to be low. Thus, due to this context dependency and low accuracy, evaluation of genetic diversity is a complex subject that demands additional studies based on phenotypic and molecular marker data to provide useful information about populations undergoing selection.

Iowa Stiff Stalk Synthetic (BSSS) is an important source of elite parental lines that have been used in the United States and other countries (Troyer, 2000). BSSS has been improved by intrapopulation and reciprocal recurrent selection beginning in 1939 and 1949, respectively (Sprague, 1946; Smith, 1983). Assessment of genetic variation in BSSS undergoing intrapopulation recurrent selection has been made based on phenotypic data (Eberhart et al. 1973; Helms et al., 1989; Lamkey, 1992; Holthaus and Lamkey, 1995). The general purpose of this study is to evaluate the effects of 14 cycles of RS, seven cycles of

half-sib following seven cycles of S_2-progeny selection, based on RFLP data, on the genetic structure of BSSS motivated by the following questions: 1) How much variation existed in the parental population? 2) Which selection method contributed the most to changes in genetic variation? 3) Was there a significant founder effect during the cycles of selection? 4) What are the predicted heterozygosity for future cycles of selection? 5) Are the populations in Hardy-Weinberg equilibrium for the RFLP loci we sampled? 6) How genetically diverse are the cycles of selection based on genetic distances and principal component analysis? 7) What were the genetic contributions of the progenitor lines to the cycles of selection? 8) Are the genetic distances among parental lines good predictors of single-cross performance? 9) Were there RFLP loci with changes in allele frequencies greater than could be explained by genetic drift alone?

Dissertation Organization

The dissertation consists of four chapters. The first chapter is the general introduction. The second and third chapters are manuscripts that will be submitted to Crop Science. The first manuscript discusses genetic diversity within cycles of selection of the intrapopulation recurrent selection program in BSSS maize population and addresses the research questions 1 to 5. The second manuscript discusses genetic diversity among cycles of selection of the intrapopulation recurrent selection program in BSSS maize population and addresses the research questions 6 to 9. The fourth chapter is the general conclusions.

Literature Review

Iowa Stiff Stalk Synthetic (BSSS)

In 1933 and 1934 breeders from diverse programs selected 16 inbred lines based on their superior performance for stalk quality (Sprague, 1946). These 16 lines were intermated in a

series of eight single-crosses, four double-crosses, and one double-double cross to form the Iowa Stiff Stalk Synthetic (Labate et al., 1999). Hallauer et al. (1983) and Messmer et al. (1991) reported that 10 of these lines were selected from the Reid Yellow Dent (RYD) population and six traced to other sources. The combining ability pattern of BSSS indicates that it can be classified as belonging to the Reid Yellow Dent heterotic group because this synthetic show higher heterosis in crosses to populations related to Lancaster Sure Crop than to populations related to Reid Yellow Dent types (Hallauer et al., 1983).

Genetic variation among the parental lines of this synthetic has been evaluated based on quantitative genetic analysis of phenotypic data (Stucker and Hallauer, 1992) and molecular data (Neuhausen, 1989; Messmer et al., 1991; Labate et al., 1997). Combining ability for grain yield and other agronomic traits was evaluated by Stucker and Hallauer (1992). They crossed each parental line with 12 other parental lines and a total of 96 crosses was obtained and distributed in 6 sets of Design II crosses that were evaluated for 2 years at 3 Iowa locations. They found considerable variation for performance of single-crosses and general and specific ability effects for grain yield. Also, they reported significant variation for general combining ability effects for other quantitative traits, such as plant and ear height, and root and stalk lodging. Based on a sample of 70 RFLP loci, Neuhausen (1989) observed that about 23% of the alleles were unique, i.e., traced exclusively to a BSSS parental line. Based on a sample of 82 RFLP loci, Labate et al. (1997) found that about 25% of the alleles of BSSS were unique and estimated the expected heterozygosity across all loci as 0.54. Also, Labate et al. (1997) reported that two lines made greater contributions than the other lines to the original and 12^{th} cycles of recurrent reciprocal selection in BSSS population. Messmer et al. (1991) evaluated the parental lines of BSSS based on 144 RFLP and 22 allozyme loci.

They also observed that a high frequency, about 25%, of the RFLP and allozyme alleles, traced exclusively to a BSSS parental line. Additionally, they reported great variation for genetic distances among the parental lines of BSSS. The results of these studies indicated that there is a considerable genetic variation among the parental lines of the Iowa Stiff Stalk Synthetic. These studies did not address the questions of whether the genetic distance among the parental lines of BSSS are good predictors of their single-crosses performance or if there are some progenitor lines with low genetic divergence to the cycles of selection of the intrapopulation recurrent selection program in BSSS.

The Recurrent Selection Program In BSSS Population

Seven cycles of HS selection followed by 9 cycles of S_2 selection for grain yield and other traits have been conducted in the BSSS population. The details of the selection program were given by Eberhart et al. (1973) and Lamkey (1992).

A double cross hybrid, IA 13, was used as the tester during the HS selection period. Sprague (1946) stated: "Iowa hybrid 13 is a high-yielding, rather weak-stalked double-cross". This hybrid, for the period of 1934 to 1939, had the best performance for grain yield in the South and Central regions of Iowa (Sprague, 1946). Principal objectives of this program were to increase grain yield and to reduce stalk and root lodging, grain moisture, plant and ear height, and frequency of dropped ears (Hallauer et al., 1983). Several studies have evaluated this program based on phenotypic data. Nonsignificant changes of additive 2%) and dominance variance components for grain yield were reported by Holthaus and Lamkey (1995) for the seven cycles of half-sib selection. Considering S_2 progeny selection, they also observed nonsignificant changes in additive and dominance variance components for grain yield. Some estimates of response for 7 HS cycles of selection for grain yield, based

on performance per se, were 1.35, 3.9, and 1.9% cycle^{-1} according to Eberhart et al. (1973), Lamkey (1992), and Holthaus and Lamkey (1995), respectively. Six cycles of S_2 progeny selection resulted in a lower average response for this trait; i.e., 1.2% per cycle; considering 13 cycles of selection (seven cycles of HS plus six cycles of S_2-progeny selection), the genetic gain was 21.4% or 0.84 Mg ha^{-1}, an average of 1.6% per cycle (Holthaus and Lamkey, 1995). A common explanation for low genetic gains for grain yield in the population for S_2-progeny selection in BSSS was inbreeding depression, due to small effective population size (Helms et al., 1989; Lamkey, 1992; Holthaus and Lamkey, 1995).

Holthaus and Lamkey (1995) also reported that ratios of additive and dominance variances for grain yield after 13 cycles of selection were 0.76 and 1.1, respectively. A group of 200 random lines was developed from BS13(S)C0 for 5 generations (S_0 to S_4) and was evaluated at five locations (Edwards, 1999). He also observed low ratios for additive and dominance variance for grain yield, i.e., 0.91 and 0.68 for noninbred and inbred individuals, respectively. These ratios are lower than an average of 1.6 estimated for other maize populations (Hallauer and Miranda, 1988), indicating that BSSS has unusual high proportion of dominance variance (Hallauer et al., 1983; Holthaus and Lamkey, 1995; Edwards, 1999). Edwards (1999) reported that grain yield was the only trait where inbred and noninbred individuals showed correlation between genotypic (G) and breeding (A) values lower than the correlation between G and dominance deviations (D). Also, inbred individuals had a lower G vs. A correlation than noninbred ones for this trait. Edwards (1999) concluded that selection among selfed progenies for grain yield may be more efficient for reducing inbreeding depression than improving the performance per se of BS13(S) population.

Genetic Structure Measures

Observed allelic and genotypic frequencies of marker loci can be used to estimate several measures related to genetic structure within and among populations and cycles of selection, such as: proportion of polymorphism loci, mean number of alleles, observed heterozygosity, expected heterozygosity, F_{IS} statistics, and genetic distances.

Proportion of polymorphic loci and mean number of alleles are usual measures of genetic diversity. A polymorphic locus shows allelic frequencies higher than could be expected in a selection-mutation balance model (Falconer and Mackay, 1996). Usually, a locus is considered polymorphic when the most common allele is at a frequency lower than 0.99 or 0.95. Nei (1987) pointed out that this variable is subject to large errors when either the number of individuals or the number of loci is small. Mean number of alleles per locus can be calculated as the total number of alleles divided by the total number of loci. Nei (1987) indicated that this measure is also very sensitive to sample size, because of the high presence of rare alleles in natural populations.

The observed mean heterozygosity frequency (H_0) over all the loci is a simple measure of variation. Each heterozygote has different alleles. For noninbred populations higher frequency of heterozygotes means higher genetic variation (Weir, 1996). The expected heterozygosity frequency (H_e), also known as gene diversity, is equivalent to the heterozygosity estimated for a randomly mating population (Nei, 1987). It also can be defined as the expected frequency of heterozygous loci in an individual sampled at random (Nei, 1987). Weir (1996) recommends its use for evaluating inbred populations more than observed heterozygosity, because variation in these populations is characterized by presence of different homozygotes. For random mating populations the values of observed and

expected heterozygosity are expected to be the same. In a pure genetic drift model, expected heterozygosity can be predicted by the equation $He_t = He_0 [1- (1/2Ne)]^t$, where t is the number of generations between cycles, Ne is the harmonic mean of effective size population, He_t is the predicted value for mean expected heterozygosity in generation t, and He_0 is the expected heterozygosity estimated in the initial generation (Nei, 1987). Thus, low effective size population, high initial genetic variation, and high number of generations are factors that contributed to great loss of genetic variation. One can observe that these factors can be common in maize populations undergoing long-term recurrent selection programs.

F_{IS} statistic is defined as $[1- (H_o / H_e)]$, where H_o and H_e are the observed and expected heterozygosity frequency in a subpopulation, respectively. It is the inbreeding coefficient of individuals relative to their subpopulation (Hartl and Clark, 1997). It is a useful measure for evaluating Hardy-Weinberg condition. When F_{IS} is not significantly different from zero, the subpopulation is basically under random mating.

Standard genetic distance (Nei, 1972) is the most traditional measures of genetic differentiation among populations (see Weir, 1996 and Nei and Kumar, 2000 reviews about genetic distances). Standard genetic distance (D) is given by:

$$D = - \ln [\Sigma_l \Sigma_i p_{1li} p_{2li} / (\Sigma_l \Sigma_i p^2_{1li} \Sigma_l \Sigma_i p^2_{2li})^{0.5}],$$

where p_{1li} and p_{2li} are the frequency of allele i of locus l in populations 1 and 2, respectively. According to Hedrick (1999), the standard genetic distance can be predicted as: $D = - 0.5 \ln [(1- He_0) / (1 - He_0 (1 – 1/2Ne))t]$. Considering genetic drift, the same factors (low effective population size, high initial genetic variation, and high number of generations) that cause loss of genetic variation within subpopulations also increase the variation among subpopulations. Thus, one can expect that genetic drift can be an important factor causing reduced genetic

variation within and increased genetic distances over generations in long-term maize recurrent selection programs.

According to Kimura (1957), when the model is additive, the probability of fixation, u(p), of an allele can be estimated by: $u(p) = (1 - e^{-2Nesp_0}) / (1 - e^{-2Nes})$, i.e., favorable alleles, with high coefficient of selection, s, can show u(p) higher than p_0, the probability of fixation is directly proportional to the initial allele frequency, and neutral and effectively neutral alleles (ones with no or weak coefficient of selection, respectively) show $u(p) \approx p_0$ (Hedrick, 2000). This equation, although based on a simple model, illustrates a very important point: the probability of fixation of an allele in a given population is dependent on the product Nes, i.e., a complex relationship between genetic drift and selection.

Evaluation of Recurrent Selection Maize Programs Based On Isozyme Markers

Isozyme data have been used to evaluate recurrent selection programs in some maize populations. As expected, the results of different studies are variable and dependent on diverse factors such as: population, breeding method, effective population size, sample loci, sample individuals, and number of cycles of selection.

Brown (1971) evaluated the effects of eight long-term Illinois selection programs for grain quality (high and low oil, high and low protein, reverse high and low oil, and reverse high and low protein populations) on the allele frequencies of six isozyme loci. The genotypic frequencies in all populations were consistent with Hardy-Weinberg ratios. He also observed that each population was distinct from others, considering all the loci evaluated. Brown (1971) concluded that most of the observed differences in allele frequencies among the eight populations could be explained by random genetic drift.

Brown and Allard (1971) evaluated two cycles of reciprocal recurrent selection (RRS) in two synthetics, A and B. In general, the genotypes frequencies followed Hardy-Weinberg expectations. The A X B interpopulation cross showed higher heterozygosity than either the A or B synthetic, following what is expected to occur in RRS programs. The nine isozyme loci evaluated were not responsive for RRS. They concluded that some observed allelic frequency changes could better be explained by genetic drift, which generated linkage disequilibrium and also loss of rare recombinants at closely linked loci.

Based on four isozyme loci, Stuber and Moll (1972) compared nine cycles of full-sib (FS) selection with 20 generations of random mating in a single cross (NC7xCI21) population. The initial frequencies of the dominant alleles at three peroxidase loci were around 0.50. These alleles increased in frequency and were fixed in both subpopulations. The initial frequency of a codominant phosphatase allele F initially was 0.47. The final frequencies of this allele were 0.6 and 0.46 in the selected and unselected subpopulations, respectively. Also, the frequency of allele F and grain yield were highly correlated over the first six cycles of selection. Based on this limited information, they raised the possibility that the phosphatase locus may be associated, directly or by hitchhiking, with grain yield in that population.

Stuber et al. (1980) evaluated changes in allelic frequencies of 20 enzyme loci after nine cycles of RRS and nine cycles of intrapopulation FS selection in the populations Jarvis and Indian Chief. Only eight loci were influenced significantly by selection; the other 12 did not show variation and were not considered for further analysis. They did not find evidence of deviations from Hardy-Weinberg proportions at the eight loci. They also observed a few cases of linkage disequilibrium among all pairs of loci evaluated. Stuber et al. (1980)

reported that the primary cause of allelic changes at the eight loci was directional selection since significant linear changes in allele frequencies occurred after rejecting the drift hypothesis of Shaffer et al. (1977) test. For the Jarvis population, the change in allelic frequencies was similar in magnitude and direction for both methods of selection. Probably, this was the first study in maize to perform a proper statistical test to evaluate changes in allele frequencies in populations undergoing selection.

Kahler (1983) monitored allelic frequencies at nine enzyme loci in two strains of the Krug BSK population, improved by eight cycles of HS and S_1 independent programs of selection. In general, changes in allelic frequencies were small and could be explained by genetic drift alone. Linear trends observed at one locus for HS selection and three loci for S_1 selection may have been caused by directional selection. According to Kahler (1983), stabilizing selection could be a possible explanation for the "fluctuation and reversals" pattern of changes that occurred for some loci.

Revilla et al. (1997) evaluated three cycles of S_1 progeny selection in two populations based on 20 isozyme loci, a number of loci well above the average used in previously reported studies. Ten progenies were selected per cycle in each population. Although the effective population size was small in both populations, three measures of genetic diversity (number of alleles per locus, number of polymorphic loci, and mean heterozygosity) did not reveal significant reduction in genetic variation among the cycles of selection for the loci evaluated.

Evaluation of Recurrent Selection Maize Programs Based On Molecular Markers

Few studies of genetic variation in maize populations under RS have been conducted based on molecular data. Heredia-Diaz et al. (1996) monitored six cycles of bi-directional

recurrent phenotypic selection (RPS) for rind penetrometer resistance in the MoSCSSS maize population, a relative of BSSS. RPS was made before flowering and 120 individuals were selected for each cycle of each subpopulation, a high effective population size. Initially, 102 RFLP loci were evaluated by comparing the two opposite sixth cycles. Only 43% of the loci showed some indication of changes in allelic frequencies. Within this group, 16 single-copy RFLP loci were studied further in the even numbered cycles of selection. A representative number of plants (90 to 176) were sampled per cycle. The authors reported that the change in pattern of allelic frequencies in eight of the 16 RFLP loci studied was caused by significant directional selection effects and could not be explained by genetic drift alone.

Labate et al. (1997, 1999, and 2000) monitored changes in allelic and genotypic frequencies of 82 RFLP loci in a RRS program involving BSSS and BSCB1 maize populations. In this program, 10 S_1 progenies were selected per population from cycle1 (C1) to C8 and 20 S_1 progenies were selected from C9 to C12. Labate et al. (1997) reported that in BSSS gene diversity was reduced from 0.49 to 0.34 after 12 cycles of selection. For BSCB1, gene diversity was 0.49 in C0 and was reduced to 0.32 in C12. However, the Nei's genetic distance between the populations increased from 0.21 to 0.66. Loss of genetic variation within and an increase in genetic distance between populations follows what is expected to occur in RRS programs.

Labate et al. (1999) observed 27 RFLP loci with allelic frequency changes greater than could be explained by genetic drift alone, indicating that 33% of loci evaluated were affected by selection. They also estimated effective population size considering two samples: one including all the alleles and other excluding rare alleles. The effective population size estimates for nonneutral and neutral loci were smaller than the harmonic mean of the number

of selected progenies (N) and close to 2N – 1, respectively. Considering all loci, the estimates of effective population size were between N and 2N - 1 and close to N for the first and second kind of sample data, respectively, indicating that samples including rare alleles can bias estimations of effective population size upwards (Labate et al., 1999).

Labate et al. (2000) and Popi et al. (2000) used principal component analysis (PCA) based on allele frequencies to evaluate reciprocal selection programs in maize. In both studies it was possible to distinguish cycles of selection based on this technique, revealing that PCA can be an alternative to evaluate genetic variation between and within cycles of selection. Labate et al. (2000) also evaluated Hardy-Weinberg (H-W) and linkage equilibrium in a RRS program involving BSSS and BSCB1 maize populations. For both populations the genotypic frequencies at the majority of loci followed Hardy-Weinberg proportions. The proportion of pairs of loci in linkage disequilibrium increased much more over the cycles of selection for BSSS than for BSCB1 population. However the authors suspected that this occurred probably due to the artificial inclusion of three BSSS(R)C0 individuals in the sample of BSSS(R)C12 population (Labate et al., 2000). As expected and similar to the results obtained with isozyme data, one can observe that the results of different studies are variable and dependent on diverse factors such as: population, breeding method, effective population size, sample loci, sample individuals, and number of cycles evaluated.

Goldman et al. (1994) evaluated the effects of the Illinois bi-directional long-term maize recurrent selection for protein concentration on kernel oil concentration and kernel weight. They evaluated S_1 progenies from the cross IHP (Illinois high protein) x ILP (Illinois low protein) based on molecular (100 RFLP loci) and phenotypic (kernel oil concentration and kernel weight) data. Variation at 25, 18, and 6 RFLP loci were significantly associated with

kernel oil, kernel weight, and both traits, respectively. Additional information could be accumulated about populations, breeding methods, traits undergoing selection, and marker loci if different programs of recurrent selection were evaluated based on phenotypic data and on a standard larger set of the same marker loci.

Comparing The Utility Of Isozymes, RFLPs, And Microsatellite Markers To Evaluate Genetic Diversity In Maize

A major disadvantage of isozyme data is the limited number of loci available to study quantitative traits. Dubreil and Charcosset (1998) compared the genetic structure of 10 maize populations based on 20 isozymes and 35 RFLP loci. They reported that RFLPs provided higher polymorphism and total diversity measure than isozymes. Messmer et al. (1991) evaluated the genetic variation of 21 maize inbreds based on 144 RFLP and 22 isozyme loci. The RFLP assay revealed higher values for expected heterozygosity and Rogers' distance. It also showed better distinction of groups by principal components analysis. They reported that RFLP provided better information related to genetic diversity than isozymes and recommended that precise estimates of genetic distances require a large number of RFLP loci. Their results suggest that RFLPs have an advantage over isozymes in studying genetic variation. However, this does not reduce the merits of the pioneering isozyme assays, that have provided a large amount of information on the genetic structure of maize and other crops. The advent of isozymes also stimulated the development of several methods for evaluating the genetic structure of populations, that are still in use today.

RFLPs are codominant markers found with high abundance in the genome; they are easy to score and interpret, although they need a large amount of genomic DNA and the development of probes are labor and time intensive (Rafalski et al., 1996). Since evaluation

of genetic structure demands estimation of genotypic frequencies, the use of codominant markers is recommended because they permit discrimination of observed heterozygotes, a variable that also is a component in calculation of inbreeding coefficients.

Other codominant markers, such as microsatellites (SSRs), can be used for evaluation of genetic structure of populations undergoing selection. Microsatellites alleles are characterized by PCR amplifications of regions with variable number of short (usually di or tri- nucleotide sequences) tandem repeats (Kochert, 1994). Studies conducted by Smith et al. (1997) and Bernardo et al. (2000) demonstrated that microsatellites are useful tools to evaluate genetic associations among inbred lines. Smith et al. (1997) compared data from pedigree, 131 SSR loci, and 80 RFLP loci to evaluate the association among 58 maize inbred lines. Pairwise inbred lines genetic distances based on RFLP and SSRs data were highly correlated (0.85). Also, both estimates of pairwise genetic distances were highly correlated with the coancestry coefficient based on pedigree data (0.80 and 0.81, respectively). Bernardo et al. (2000) compared data from pedigree, 195 SSR loci, and 124 RFLP loci to evaluate the association among 13 maize inbred lines. Estimates of coefficient of coancestry based on different kind of data were all highly correlated; i.e., 0.87, 0.92, and 0.97 for SSR vs. RFLP, SSR vs. pedigree, and RFLP vs. pedigree comparisons, respectively. According to Smith et al. (1997) microsatellites are better molecular markers for maize, because they are cheap, reliable, less time intensive, and, for a given sample of SSR loci, the genetic variation within and among diverse populations can be easily estimated and interpreted. Although microsatellites show great potential to be used in studies of evaluation of maize recurrent selection programs, there are no reports of the utility of microsatellites for genomic evaluation of maize populations undergoing selection.

References

Bernardo, R., J. Romero-Severson, J. Ziegle, J. Hauser, L. Joe, G. Hookstra, and R. W. Doerge. 2000. Parental contribution and coefficient of coancestry among maize inbreds: pedigree, RFLP, and SSR data. Theor. Appl. Genet. 100: 552-556

Brown, A.H.D. 1971. Isozyme variation under selection in *Zea mays*. Nature 232: 570-571.

Brown, A.H.D., and R.W. Allard. 1971. Effect of reciprocal recurrent selection for yield on isozyme polymorphisms in maize (*Zea mays L.).* Crop Sic. 11: 888-893.

Dubreil, P., and A. Charcosset. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor. Appl. Genet. 96: 577-587.

Eberhart, S.A., S. Debela, and A.R. Hallauer. 1973. Reciprocal recurrent selection in the BSSS and BSCB1 maize populations and half-sib selection in BSSS. Crop Sci. 13: 451-456.

Edwards, J.W. 1999. Observed and predicted phenotypic effects of inbreeding in the BS13(S)C0 maize population. Ph.D. thesis, Iowa State University, Ames, Iowa.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, England.

Goldman, I.L., T.R. Rocheford, and J.D. Dudley. 1994. Molecular markers associated with maize kernel oil concentration in an Illinois high-protein x Illinois low-protein cross. Crop Sci. 34: 980-915.

Hallauer, A.R., and J.B Miranda Fo. 1988. Quantitative genetics in maize breeding. 2nd ed. Iowa State Press, Ames, Iowa.

Hallauer, A.R., W.A. Russell, and O.S. Smith. 1983. Quantitative analysis of Iowa Stiff Stalk Synthetic. p. 83-104. In: J.P. Gustafson (ed.) Stadler genetics symposium, 15th, Columbia, MO

Hartl, D.L., and A.G. Clark. 1997. Principles of population genetics. 3rd ed. Sinauer, Massachusetts.

Hedrick, P.W. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313-318.

Hedrick, P.W. 2000. Genetics of populations. 2nd ed. Jones & Bartlett, Massachusetts.

Helms, T.C., A.R. Hallauer, and O.S. Smith. 1989. Genetic drift and selection evaluated from recurrent selection programs in maize. Crop Sci. 29: 602- 607.

Heredia-Diaz, O., A. Alsirt, , L.L. Darrah, and E.H. Coe. 1996. Allelic frequency changes in the MoSCSSS maize synthetic in response to bi-directional recurrent selection for rind penetrometer resistance. Maydica 41: 65-76.

Holthaus, J.H., and K.R. Lamkey. 1995. Populations means and genetic variances in selected and unselected Iowa Stiff Stalk Synthetic maize populations. Crop Sci. 35: 1581-1589.

Kahler, A.L. 1983. Effect of half-sib and S1 selection for increased grain yield on allozyme polymorphism in maize. Crop Sci. 23: 572- 576.

Kochert, G. 1994. RFLP technology. p. 8-38. In R.L. Phillips and J.K. Vasil (eds.). DNA-based markers in plants. Kluwer Academic Publishers, Netherlands.

Kimura, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28: 882-901.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman.1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416-423.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1999. Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. 99: 1166-1178.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 2000. Hardy- Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45: 243-255.

Lamkey, K.R. 1992. Fifty years of recurrent selection in the Iowa Stiff Stalk Synthetic maize populations. Maydica 37: 19-28.

Lynch, M. 1995. A quantitative-genetics perspective on conservation issues. p. 471-494. In J.C. Avise and J.L. Hamrick (eds.). Conservation genetics: case histories from nature. Chapman & Hall, New York, New York.

Messmer, M.M., A.E. Melchinger, M. Lee, W.L. Woodman, E.A. Lee, and K.R. Lamkey. 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize populations: comparison of allozyme and RFLP data. Theor. Appl. Genet. 83: 97-107.

Nei, M. 1972. Genetic distance between populations. Amer. Nat. 106: 283-292.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York.

Neuhausen, S.L. 1989. A survey of Iowa Stiff Stalk Synthetic parents, derived inbreds, and BSSS(HT)C5 using RFLP analysis. Maize Gen. Coop. Newsl. 63: 110-111.

Popi, J., L. Rajnpreht, L.W. Kannenberg, and K.P. Pauls. 2000. Random amplified polymorphic DNA-based evaluation of diversity in the hierarchical, open-ended population enrichment maize breeding system. Crop Sci. 40: 619-625.

Rafalski, J.A, J.M. Voguel, M. Morgante, W. Powell, A. Chaz, and S.V. Tingey. 1996. Generating and using DNA markers in plants. p. 75 – 134. In B. Birren and E. Lay (eds.). Nonmammalian genomic analysis: A practical guide. Academic Press.

Revilla, P., M.L. Vales, R.A. Malvar, and A. Ordas. 1997. Allozyme frequencies, heterozygosity and genetic distances following S_1 recurrent selection in two synthetic maize populations. Theor. Appl. Genet. 95: 1057-1061.

Russell, P.J. 1996. Genetics. Sinauer. 4th ed. New York.

Smith, J. S. C., E. C. L. Chin, H. Shu, O. S. Smith, S. J. Wall, M. L. Senior, S. E. Mitchell, S. Kresovich, and J. Ziegle. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays L.*): comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95:163-173

Smith, O.S. 1983. Evaluation of recurrent selection in BSSS, BSCB1, and BS13 maize populations. Crop Sci. 23: 35-40.

Sprague, G.F. 1946. Early testing of inbred lines of corn. J. Am. Soc. Agron. 38: 108-117.

Schaffer, H.E., D.G. Yardley, and W.W. Anderson. 1997. Drift or selection: a statistical test of gene frequency variation across generations. Genetics 87: 371-379.

Stuber, C.W., and R.H Moll. 1972. Frequency changes of isozymes alleles in a selection experiment for grain yield in maize (*Zea mays L.*). Crop Sci. 12: 337-340.

Stuber, C.W., R.H Moll, M.M.Goodman, H.E. Schaffer, and B.S. Weir. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays L.*). Genetics 95: 225-237.

Stucker, D.S. and A.R. Hallauer. 1992. Genetic variability as affected by selection in Iowa Stiff Stalk Synthetic maize. Journal of Heredity 83: 410-418.

Troyer, F. 2000. Temperate corn - background, behavior and breeding. p. 393 - 463. In Hallauer, A.R. (ed). Specialty corns. 2nd ed. CRC Press Inc., Boca Raton, USA.

Weir, B.S. 1996. Genetic data analysis II. 2nd ed. Sinauer, Massachusetts.

CHANGE RFLP ALLELE AND GENOTYPIC FREQUENCIES AFTER HALF-SIB AND S_2 SELECTION IN THE IOWA STIFF STALK SYNTHETIC MAIZE (*Zea mays L.*) POPULATION

A paper to be submitted to Crop Science

Paulo Evaristo de Oliveira Guimarães and Kendall Raye Lamkey

Abstract

Evaluating of Recurrent Selection (RS) programs can increase our knowledge of methods, populations, and traits and lead to better management of breeding programs. The objective of this research was to evaluate the effects of seven cycles of half-sib selection followed by seven cycles of S_2 selection on the genetic structure of Iowa Stiff Stalk Synthetic (BSSS) maize population. BSSSP (16 progenitor lines), BS13(S)C0 (90 individuals from C7 of HS selection and C0 from S_2 selection), and BS13(S)C7 (102 individuals from 7^{th} S_2 cycle of selection) cycles were genotyped with 105 RFLP loci. The expected heterozygosity, total number of alleles, and average frequency of the most common allele were 0.58, 439, and 0.56 for BSSSP; 0.34, 251, and 0.75 for BS13(S)C0; and 0.26, 231, and 0.82 for BS13(S)C7, respectively. These, and results from other measures of genetic diversity, suggest that: (i) BSSSP is a synthetic with significant variation; (ii) the greatest reduction in diversity occurred during the half-sib selection; (iii) low average effective population size was an important factor in loss of genetic variation. The founder effect observed herein may partially explain reduced genetic gains during the S_2 selection reported in other studies. Limited RFLP diversity in BS13(S)C7 suggests that this population may not have enough genetic variability to sustain long-term genetic gains.

Introduction

Recurrent selection (RS) is a cyclic process of developing, selecting, and recombining individuals. In general, RS programs are characterized by intense selection pressure and low effective population size. For this reason, selection and genetic drift are considered the principal causes of loss of genetic variation in populations undergoing RS. The maor challenge for long-term RS programs is to manage the genetic variability to sustain short, medium, and long-term genetic gains.

Iowa Stiff Stalk Synthetic (BSSS) has been improved by intrapopulation and reciprocal RS beginning in 1939 and 1949, respectively (Sprague, 1946; Smith, 1983). BSSS is a source of public and private parental elite lines that have been used in the United States and other countries (Troyer, 2000). The intrapopulation RS program in BSSS has been characterized by seven cycles of half-sib (HS) selection followed by nine cycles of S_2-progeny selection. Principal objectives of this program were to increase grain yield and to reduce stalk and root lodging, grain moisture, plant and ear height, and frequency of dropped ears (Hallauer et al., 1983). Quantitative genetic analysis of BSSS phenotypic data has produced vast amounts of genetic information, including genetic component of variances, heritability, combining ability, inbreeding depression, and responses to selection (see review in Hallauer and Miranda, 1988). The most recent studies of response in the populations per se for grain yield revealed that the average genetic gain during the HS selection was greater than that observed for the first six cycles of S_2-progeny selection (Lamkey, 1992; Holthaus and Lamkey, 1995). Genetic gain for grain yield was less than expected for S_2-progeny selection and raises concern about the potential for long-term selection responses in the population per se.

Molecular markers can provide information about the effect of recurrent selection on the genetic variation of populations. Polymorphism level, number of alleles, and observed and expected heterozygosity are examples of measures of genetic variation that can be generated by molecular marker data (Nei, 1987). There have been a few studies where molecular markers have been used to study genetic variation in maize populations undergoing RS (Heredia-Diaz et al., 1996, Labate et al., 1997, 1999, and 2000; Popi et al., 2000). Labate et al. (1997) monitored changes in allelic and genotypic frequencies of 82 RFLP loci in a Reciprocal Recurrent Selection (RRS) program involving the BSSS and BSCB1 maize populations. They reported that after 12 cycles of selection diversity was reduced within both populations but the genetic distance between them increased, following what was expected to occur in RRS programs.

Development of experiments that focus on changes in genetic structure of populations undergoing RS can lead to the accumulation of additional knowledge of RS methods, populations, and traits. This knowledge can be used to determine more efficient ways to maintain genetic gains in long-term selection programs. Although the intrapopulation RS program in BSSS population has been extensively evaluated phenotypically, there has been little evaluation of this program using molecular marker data. The objective of our study was to evaluate the effects of RS on the genetic structure of BSSS by using RFLP data. Our goal was to obtain data to answer the following questions: (i) How much variation existed in the parental population? (ii) Which selection method contributed the most to changes in genetic variation? (iii) Was there a significant founder effect during the cycles of selection? (iv) What are the predicted heterozygosities for future cycles of selection? (v) Are the populations in Hardy-Weinberg equilibrium for the RFLP loci we sampled?

Materials and Methods

Genetic Material

BSSS is a synthetic formed by intermating 16 inbred lines previously selected for stalk quality (Sprague, 1946). Seven cycles of HS selection followed by nine cycles of S_2 selection for grain yield and other traits has been conducted in the intrapopulation RS program. A double cross hybrid, IA 13, was used as the tester in the HS selection program. The details of the selection program were given by Lamkey (1992).

The harmonic mean of the number of selected progenies was 10 for HS and 15 during the six cycles of S_2 selection. Intensities of selection averaged 9.7 and 15.6% for HS and S_2 selection, respectively (Holthaus and Lamkey, 1995). This RS program was genotyped with molecular markers by sampling three populations: BSSSP (collection of the 16 parental lines), BS13(S)C0 (seventh cycle of HS selection and C0 of S_2-progeny selection), and BS13(S)C7 (seventh cycle of S_2-progeny selection). The collection of progenitor lines included 14 of the 16 original lines plus the parental lines (Fe and IndB2) of line F1B1 (Labate et al., 1999). The line CI617 and its parental lines have never been found. Ten of these lines were selected from the Reid Yellow Dent population and six lines were developed from other populations (Messmer et al., 1991). Syn-2 generations from BS13(S)C0 and BS13(S)C7 populations were obtained by random mating about 350 individuals from the Syn-1 populations. For our study, 90 and 102 individuals from the Syn-2 generation were used to represent BS13(S)C0 and BS13(S)C7 populations, respectively.

RFLP Analysis

Sampling of the populations followed the approach used by Labate et al. (1997). Progenitors lines and the populations were grown in the same environment. Mature leaf

tissue from randomly chosen individuals were collected, freeze-dried, ground, and stored at -20^{0}C. DNA isolation and RFLP analysis were performed at Biogenetic Service INC that also provided an initial summarization of the observed fragment data. DNA was isolated using a CTAB extraction method. Techniques for DNA isolation and other RFLP methods are found in Ausubel (1994) with slight modifications by Biogenetic Services, Inc. (unpublished). DNA was digested using two restriction enzymes, Hind III and EcoR1, depending on the probe employed.120 Gel electrophoresis was conducted with a TAE buffer. Southern blots were made on nylon membranes and the resulting complex of fragments was subject to hybridization techniques (Budowle and Baechtel, 1990). After washing the membranes to remove the unhybridized probe solution, they were placed on Kodak AR X-ray film and exposed for up to 14 days depending on the optimal exposure time. Genotypes were collected for 100 probe-enzyme combinations. Eighty-one and 19 of these combinations were made with RFLPs obtained by using restriction enzymes Hind III and EcoR1, respectively. The RFLP probes are distributed on all 10 maize chromosomes. Information about origin, sequence, location, and references about these probes can be obtained at Maize Genetic Database (http:// www.agron.missouri.edu). A molecular weight was assigned to each band by using Lambda markers. Each band was considered to be an allele. Some alleles, however, were associated with more that one band. Probe BNL05.24 was discarded because the banding patterns were too complex to be interpreted. Six probes (BNL05.14, BNL13.05, UMC042, UMC055, UMC 065, and UMC158) mapped to two loci. A total of 105 RFLP loci were detected (Table 1). Data were scored as presence/absence of each band and stored in a database developed by Frisch (M. Frisch, personal communication, 2000)

Statistical Analysis

Observed allelic and genotypic frequencies for each RFLP locus were used to estimate the following statistics related to genetic structure among and within cycles of selection: proportion of polymorphic loci, frequency of the most common allele, number of alleles, observed heterozygosity, expected heterozygosity, Hardy-Weinberg equilibrium, and F_{IS} statistics.

A locus was considered as polymorphic when the frequency of the most common allele was less than 0.95. For each locus, allelic frequency was estimated as the observed count of an allele divided by the total number of alleles. For each population, the most common allele was identified for each locus. Mean number of alleles was estimated as the total number of alleles divided by the total number of loci.

Observed heterozygosity for each locus, $\overline{H}_{0l}$, was estimated as the count of all heterozygotes divided by the total number of sampled genotypes. Observed heterozygosity averaged over all loci for each population, $\overline{H}_{0}$, was estimated as the mean of observed heterozygosity across all sampled loci. Expected heterozygosity by locus, $\overline{H}_{el}$, was estimated by using the unbiased formula of Nei (1978). This measure, also known as gene diversity, is equivalent to the heterozygosity estimated for a randomly mating population (Nei, 1987). Average expected heterozygosity per population, $\overline{H}_{e}$, was estimated as the mean of expected heterozygosity across all sampled loci. The sampling variance $V(\overline{H}_{e})$, was estimated as $\Sigma_l^n(\overline{H}_{el} - \overline{H}_{e})^2 / n(n-1)$ for n loci (Nei and Kumar, 2000). Since the cycles of selection are closely related, pairwise t-tests were used to compare the differences in $\overline{H}_{e}$ among them (see Nei and Kumar, 2000).

A locus is in Hardy-Weinberg (H-W) equilibrium when the genotypic frequencies are the products of allele frequencies (Weir, 1996). Exact tests were used for checking departures from H-W proportions (Weir, 1996). For each locus, the exact significance level for H-W proportions was determined in 3200 samples obtained by a shuffling process (Lewis and Zaykin, 2001). Either an excess or a deficiency of heterozygotes was estimated by the F_{IS} statistic. F_{IS} is defined as $[1-(\overline{H}_0/\overline{H}_e)]$ and is the average inbreeding coefficient of a subpopulation (Hartl and Clark, 1997). Confidence intervals (95%) for F_{IS} were obtained by bootstrap analysis (1000 replicates) over all loci. The Genetic Data Analyses software package (Lewis and Zaykin, 2001) was used for computing H-W tests, F_{IS} statistics, proportion of polymorphic loci, allele frequencies, number of alleles, and observed and expected heterozygosity.

Mean expected heterozygosity values were also predicted for BS13(S)C0 and BS13(S)C7 populations based in the equation $He_t = He_0 [1- (1/2Ne)]^t$ provided by Nei (1987), where t is the number of generations between cycles, Ne is the harmonic mean of effective size population, He_t is the predicted value for mean expected heterozygosity in generation t, and He_0 is the expected heterozygosity estimated in generation zero. He_{C0} and He_{C7} were predicted as $He_P [1- (1/20)]^7$ and $He_{C0} [1- (1/30)]^7$, respectively.

Results and Discussion

Changes In Number And Frequency Of Alleles

The distribution of allele frequencies was skewed towards alleles with lower frequencies for BSSSP (Fig. 1). Nearly 50% of the alleles observed had frequencies between 0.062 and 0.125, meaning that these alleles were observed in only one or two progenitor lines. The distributions for C0 and C7 tended to a U-shape, i.e., higher frequencies for extreme rather

than for intermediate allele frequencies. BSSSP had a lower mean frequency of alleles than C0 and C7 (Table 2).

Proportion of polymorphic loci (most common allele frequency less than 0.95) and number of alleles decreased over the cycles of selection (Table 2). Only two loci were monomorphic for BSSSP in contrast with about 20 and 33% monomorphic loci observed in C0 and C7, respectively. Compared to the BSSSP population, the total number of alleles decreased 43% in C0 and 47% in C7. The mean number of alleles per locus was lower in C0 and C7 than in BSSSP (Table 2). About 40% of the alleles in BSSSP were unique, i.e., not detected in C0 and C7 sampled populations (Table 2). The average frequency of the unique alleles in P was 0.11 indicating that most of the alleles not detected in C0 and C7 had low frequencies in BSSSP. This observation is consistent with what is expected under random genetic drift, where rare alleles have higher probabilities of being lost. The probability of losing a neutral allele is approximately equal to $1 - p_0$, where p_0 is the initial frequency of the allele (Comstock, 1996).

These data reveal a significant reduction in the sampled genetic variability over the cycles of selection (Fig. 1 and Table 2). The genetic variability detected in C0 was similar to C7, rather than being intermediate between P and C7. A possible explanation for this finding is that the 10 progenies selected per cycle during the seven cycles of HS selection caused a bottleneck effect that reduced the sampled genetic variability in C0, the founder population for the seven cycles of S_2 selection. Changes in allele frequency are directly proportional to the genetic variability and inversely proportional to the size of a base population (Hartl and Clark, 1997). Thus, greater changes in genetic variation during the HS selection period than in the following seven cycles of S_2 selection can be explained by the observation that BSSSP

had greater genetic variability than C0, which showed a high proportion of alleles at extreme frequencies.

There were five and seven unique alleles detected for C0 and C7, respectively. Migration, mutation, or miss-scoring may be the causes of the presence of unique alleles in these populations. Also some of the unique alleles in C0 and C7 may have come from the parental line CI617, which was not represented in the BSSSP population. Labate et al. (1997) also reported the presence of unique RFLP alleles in the BSSS and BSCB1 populations undergoing reciprocal recurrent selection. In our study and the study of Labate et al. (1997), the frequency of unique alleles in populations that had undergone selection was low, suggesting that contamination (migration) during recombination and maintenance of the populations was not a significant source of new alleles.

Different measures of genetic variation indicated that BSSSP has a significant amount of genetic variation (Tables 2, 3, and 4). Half of the alleles in BSSSP traced to one or two progenitor lines. Of the 439 alleles present in this synthetic, 103 came exclusively from the set of 10 lines that traced to the Reid Yellow Dent population and 104 were unique to the other six lines developed from other germplasm sources. Previous studies also found great genetic diversity and a high proportion of rare alleles in the BSSSP population (Neuhausen, 1989; Messmer et al., 1991; Labate et al., 1997).

Changes In Genotypic Frequencies.

Mean expected heterozygosity (H_e) decreased 41, 55, and 23% between BSSSP and C0, BSSSP and C7, and C0 and C7 cycles, respectively (Table 3). All three comparisons were significant ($p<0.001$) based on pairwise t-tests.

Observed and predicted H_e values were in close agreement for C7 (Table 3). For C0, however, the predicted value was larger than the observed H_e. This result is an indication that effective population size during HS selection period was lower than the harmonic mean of selected progenies. For two populations under RRS, effective size population lower than the harmonic mean of selected progenies was observed at RFLP loci where the hypothesis of pure genetic drift was rejected (Labate et al., 1999). Thus, it seems that during the HS selection period some RFLP loci were affected by selection, directly or by hitchhiking. There was variation among loci for expected heterozygosity (H_e) within and among populations (Fig. 2). For BSSSP, the distribution was skewed towards loci with high H_e values. The loci with the highest H_e values (from 0.7 to 0.9) averaged 5.8 alleles per locus compared to the average of 4.2 alleles per locus. For C0 and C7, the distribution of H_e showed a large number of loci having extremely low values (< 0.1).

There also was significant variation among individuals for observed heterozygosity (H_0) within and among populations (Fig. 3). As expected, all the highly inbred parental lines had extremely low H_0 values. Distributions for C0 and C7 indicated significant variation for H_0 within these populations and also indicated that is possible to select individuals with higher H_0 than the average of these populations. The observed variation in heterozygosity, either at a locus or on an individual basis, illustrates the importance of sample size for estimating H_0. Nei (1987) suggested that sampling a large number of loci is more important than sampling a large number of individuals for H_e estimation. He also recognized, however, the importance of sampling a large number of individuals for estimating other statistics such as allele frequency.

H_e is a traditional measure of genetic variation within populations (Weir, 1996; Hedrick, 2000). There was a significant loss of genetic variation over cycles of selection for the RFLP loci we sampled. These results are consistent with those of Labate et al. (1997) who also reported a decrease in observed and expected heterozygosity for RFLP loci in a BSSS population undergoing reciprocal recurrent selection. Analogous to what was observed for other measures of genetic diversity, H_e values were closer between C0 and C7 than between P and C0. The equation presented by Nei (1987) indicates that the magnitudes of a decrease in expected heterozygosity during a certain period is directly proportional to the initial expected heterozygosity (variation in the founder population) and inversely proportional to the effective population size (see equation on footnotes of Table 3). Thus, the small number of progenies selected in each cycle was an important factor in reducing expected heterozygosity over cycles of selection. Also, the large reduction in variability that occurred during the HS selection follows what we could expect because HS selection was characterized by a smaller effective population size and had a founder population with more variability than S_2-selection.

Holthaus and Lamkey (1995) reported that the additive genetic variance (Va) for grain yield in BSSSC0, BS13(S)C0, and BS13(S)C6 was 0.319, 0.218, and 0.228 $(Mg\ ha^{-1})^2$, respectively. Changes in Va over generations within a subpopulation caused by genetic drift can be predicted as: $Va_t = Va_o\ [1- (1/2Ne)]^t$, where Va_t is the additive variance in generation t and Va_o is the additive variance in generation 0 (Wright, 1951; Crow and Kimura, 1970). This equation was derived for traits that are predominantly additive and not acting on fitness (Hedrick, 2000). Although grain yield is a trait related to fitness, this equation was used due to absence of other alternatives to predict Va. Considering Ne = 10 and t = 7, Va for

BS13(S)C0 was predicted as 0.223 Mg ha^{-1}, and for BS13(S)C6 Va was predicted (Ne = 15 and t = 6) as 0.178 Mg ha^{-1}. There was agreement between the results obtained by RFLP and phenotypic data, because the reduction in Va for grain yield was less during S_2-selection than for HS selection, and the predicted Va was in close agreement to observed values. Since the term $[1- (1/2Ne)]^t$ is present in both equations and is related to genetic drift one may infer that random genetic drift was a major factor in reducing the expected heterozygosity (RFLP data) and additive variance (phenotypic data for grain yield)) over the cycles of selection. The equation for predicting Va, however, did not give predictions (data not shown) similar to those observed by Holthaus and Lamkey (1995) for others traits undergoing selection, i.e., grain moisture, root lodging, and ear height. Moreover, as is typical for estimates of variance components, observed differences among Va values for grain yield were not significant. Thus, results based on observed and predicted values of Va need to be considered with caution.

Previous studies have reported lower genetic gains for grain yield in the population for S_2-progeny selection in BSSS; inbreeding, due to small effective population size, was a common explanation for these results (Helms et al., 1989; Lamkey, 1992; Holthaus and Lamkey, 1995). Our results corroborate these explanations for low genetic gains in S_2-progeny selection. That is, the greatest reduction in variation at RFLP loci, assumed as neutral, occurred during the HS selection period and caused a significant founder effect for the subsequent cycles of S_2-progeny selection. Quantitative traits are expected to have a high proportion of genes with small effects. It is probable that many genes affecting grain yield and other traits in BSSS are effectively neutral. One may infer that a similar bottleneck also occurred for effectively neutral loci because they are more subject to genetic drift forces than

loci undergoing selection.. Edwards (1999) developed an additional explanation for the lack of response for grain yield during the S_2-selection period. He evaluated a group of 200 random lines was developed from BS13(S)C0 for five generations (S_0 to S_4) and was evaluated at five locations. For grain yield, correlation between genotypic (G) values and breeding (A) values for the inbred and noninbred individuals was less than the correlation between G and dominance deviations (D). In addition, for inbred individuals the correlation between G and A was smaller than the correlation between G and A for noninbred individuals. Edwards (1999) concluded that for grain yield selection among selfed progenies may be more effective for reducing inbreeding depression than for improving the performance of the population per se.

Discussion about the number of effectively neutral loci is relevant in a breeding context. As the number of genes controlling a trait increase, the number of genes with low coefficients of selection also increases (Hartl and Clark, 1997). Coefficient of selection (s) can be estimated as $s \approx i2\alpha/\sigma p$, where i is the intensity of selection, 2α is the difference between two homozygotes at a locus, and σp is the phenotypic standard deviation for the trait evaluated (Falconer and Mackay, 1996). Alleles with $s << 1/Ne$ have probabilities of fixation close to their initial allele frequency (Hedrick, 2000). Thus, a locus is effectively neutral when $2\alpha << \sigma p/Nei$, indicating that decisions about intensity of selection, effective population size, and design of experiments can affect the proportion of effectively neutral loci for quantitative trait(s) undergoing a cycle of selection. To illustrate this point, we will use the estimate of σp for grain yield of 0.606 Mg ha^{-1} for BS13(S)C6 obtained by Holthaus and Lamkey (1995) and assume Ne = 20 and i = 1.627 (Lamkey, 1992). Based on these components, $\sigma p/Nei$ was estimated as 0.019 Mg ha^{-1} for grain yield, a value that is about

0.4% of the general mean. Considering that grain yield in maize is controlled by a large number of genes, these results suggest that many loci controlling grain yield are effectively neutral. Monitoring $\sigma p/Ne_i$ over cycles of selection could be a common procedure in plant breeding for traits governed by a large number of genes. It may provide additional information that can lead to adjusting the intensity of selection, effective population size, and control of environments with the goals of increasing the chance that genes with small effects are selected.

Changes In The Frequency Of The Most Common Allele

The distribution of the most common allele (MCA) in BSSSP was concentrated around frequencies of 0.4 to 0.7, while distributions for C0 and C7 were skewed toward greater frequencies of the MCA (Fig. 4). Average frequency of MCA increased over the cycles of selection and followed what was observed for other measures of genetic diversity, i.e., greatest changes occurred during the HS selection (Fig. 4). We identified the most common allele (MCA) at each locus in each population. Thus, some loci had the same MCA while other loci had different MCA between the populations. There was more coincident of the MCA between C0 and C7 than between either P and C0 or P and C7 (Table 4). The frequencies of the MCA were larger for the loci for which the populations shared a MCA than for the loci for which populations did not share a MCA. According to Kimura (1957), when the model is additive, the probability of fixation, u(p), of an allele can be estimated by: $u(p) = (1 - e^{-2Nesp0}) / (1 - e^{-2Nes})$, i.e., favorable alleles can show u(p) higher than p_0, the probability of fixation is directly proportional to the initial allele frequency, and neutral and effectively neutral alleles show $u(p) \approx p_0$ (Hedrick, 2000). Thus, the observed increase in average frequency and higher level of coincidence for the most common allele over the

cycles of selection may be explained by genetic drift and/or selection favoring these alleles. The MCA average frequency was 0.82 for C7 (Fig. 4), indicating narrow genetic diversity for this population. Only directional selection for alleles at low frequencies or migration would increase significantly the average heterozygosity in this sample of RFLP loci.

Populations with large differences in the frequencies of the MCA at a locus are also populations that have large differences in genetic variation at that locus. Also, populations with large differences in frequencies of the MCA across loci are populations with large differences for average genetic diversity at that set of loci. This rationale can also be supported by the fact that the frequency of the MCA was highly correlated with expected heterozygosity and number of alleles (Table 5). Moreover, the average correlation of MCA with expected heterozygosity, considered as an standard measure of variation, was near by one and significantly higher than the correlation of number of alleles with expected heterozygosity (Table 5). These results, together with the utility of MCA for estimating time and probability of fixation, suggest that this variable could be used more frequently as measure of genetic diversity, although textbooks of Nei (1987) and Hedrick (2000) do not mention it among the alternatives to estimate genetic variation.

Deviations From Hardy-Weinberg Proportions

Observed F_{IS} was near to one for BSSSP and nonsignificant and close to zero for C0 and C7 (Table 3). These findings are consistent with expectations, because BSSSP is a group of 16 highly inbred lines and the samples of CO and C7 were obtained after two generations of random mating. Thus, observed and expected heterozygosity estimates were in close agreement in cycles C0 and C7 for most of the loci. Weir (1996) discussed levels of significance in multiple tests and indicated that Bonferroni correction for H-W tests is not

necessary when the primary interest is to evaluate each locus separately. Thus, we considered the 5% level of significance as appropriate for rejecting the null hypothesis of H-W equilibrium. As expected, all loci in P were not in H-W equilibrium (data not shown). Exact tests revealed 10 loci in CO and seven in C7 that showed significant deviations from H-W proportions (Table 6). Fourteen of the loci were positive for F_{IS} (Table 6), suggesting that the main reason for deviations from Hardy-Weinberg proportions was an excess of homozygotes, in agreement with previous studies in maize (Dubreuil and Charcosset, 1998; Labate et al., 2000). Some of the loci with significant H-W departure and excess of homozygosity may be directly involved in flowering time or closed linked to other loci controlling this trait, because it is usual to make pollination among plants with similar flowering times. Positive assortative mating among individuals with similar flowering time could have occurred during the multiplication of cycles C0 and C7. Other explanation could be migration, caused by contamination with pollen or seed. Positive assortative mating was also indicated in other maize studies as a probable cause of observed excess of homozygosity (Dubreuil and Charcosset, 1998; Labate et al., 2000). Only three loci with significant departures from H-W ratios were negative for F_{IS}, although two of then had F_{IS} so close to zero that they cannot be considered as having excess heterozygosity. Therefore, our results suggested that only locus asg062 in C0 population presented excess of heterozygosity. This locus also was the only one that was significant for H-W tests in both populations, although in C7 the F_{IS} value was close to zero. Thus, it seems that this locus in both populations was subject to forces that caused deviations from H-W conditions. Some possible explanations for excess of heterozygosity could be: miss-scoring, selection against the homozygotes, and contamination with pollen or seed.

Genetic Variability In The Next Cycles of Selection

Is there enough variability in BS13(S)C7 to sustain long-term genetic gains? This is a common and pertinent question raised in recurrent selection programs that requires accumulation of information for a proper answer. Within the context of this question, using a sample of RFLP loci to predict expected heterozygosity may give some clues about the potential of C7 as a founder for subsequent cycles of selection. Using again the equation $He_t = He_0 [1- (1/2Ne)]^t$, and considering Ne = 20 and $He_0 =$ 0.27, the predicted He for C9, C11, and C13 is 0.26, 0.24, and 0.23, respectively. Changing Ne to 40 and maintaining the same He_0 gave 0.26, 0.26, and 0.25 as predicted values for C9, C11, and C13, respectively. Thus, for both effective population sizes considered and for the RFLP loci that we sampled, the next six cycles of selection are predicted to show low mean expected heterozygosity values that also will be close to the heterozygosity estimated for C7. The low response for grain yield in the first six cycles of S_2-selection (Lamkey, 1992; Holthaus and Lamkey, 1995) and also the limited RFLP diversity that we estimated for BS13(S)C7 and predicted for the next cycles of selection are indicators that BS13(S)C7 may not have enough genetic variability to sustain substantial long-term genetic gains per se for grain yield. Further evaluations of the next cycles of selection in this population, based on phenotypic and molecular data, are necessary to provide additional information about the potential of this population to respond for grain yield.

Hallauer and Miranda (1988) discussed alternative methods for increasing genetic gains in recurrent selection programs. Increasing selecting pressure and using an effective population size greater than or equal to 20, and restricting the geographic target region for evaluation and selection may be options for increasing selection response in BS13(S). Labate

et al. (1997), suggested introducing variability from the parental populations as a way of increasing variability for the subsequent cycles of selection. Russell (1979) developed a new population, BS17, based on six improved versions of BSSS population. Following this idea of controlled migration with individuals of the same population, one option would be to obtain progenies from the latest cycle of BSSS(R) and/or previous cycles of selection in BS13(S) and evaluate them together with the progenies of the latest cycle of BS13(S). Progenies with superior performance and similar phenotype to BS13(S) would be selected and recombined in a way that assures a higher proportion (e.g. 75%) of selected progenies come from the latest cycle of BS13(S) cycle than from other BSSS subpopulations. With controlled migration it should be possible to increase the variability and maintain the phenotypic identity of BS13(S). Another option would be to perform few cycles of reciprocal recurrent selection inside BS13(S), or between the first and the later cycles of S_2 selection, or between BS13 and BSSSR. The subpopulations could then be merged to form a new BS13(S) base population with higher frequency of heterozygotes. Another way to manage the genetic variability in BS13 would be evaluate a large number of progenies (e.g., 500) for one cycle of selection, select a subset of progenies (e.g., 100) based on their phenotypic performance, and, based on molecular markers for the traits of interest, select a subgroup of superior progenies (e.g., 50) that results in the highest possible mean expected heterozigosity for marker loci linked to QTLs. The rationale of this approach is using molecular and phenotypic data to obtain a population with high genetic variability and superior performance that could be used as a base population for further cycles of selection. It is important to emphasize here that if the primary objective of selection in BS13(S) is to improve it as a

source of inbred lines, then per se response is not as important and modifications to the breeding system should place more emphasis on testcross response than per se response.

References

Ausubel, F (ed).1994. Current protocols in molecular biology. John Wiley and Sons, Inc.

Budowle, B., and Baehtel, FS. 1990. Modifications to improve the effectiveness of restriction fragment polymorphism typing. Appl Theor Electrophoresis 1: 181-187

Comstock, R.E. 1996. Quantitative genetics with special reference to plant and animal breeding. Iowa State University Press, Ames, Iowa.

Crow, J.F. and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Dubreil, P., and A. Charcosset. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor. Appl. Genet 96: 577-587

Edwards, J.W. 1999. Observed and predicted phenotypic effects of inbreeding in the BS13(S)C0 maize population. Ph.D. thesis, Iowa State University, Ames, Iowa.

Falconer, D.S., and T.F.C. Mackay.1996. Introduction to quantitative genetics. 4th ed. Longman, England.

Hallauer, A.R., and J.B. Miranda FO. 1988. Quantitative genetics in maize breeding. 2nd ed. Iowa State Press, Ames, Iowa.

Hallauer, A.R., W.A. Russell, and O.S. Smith. 1983. Quantitative analysis of Iowa Stiff Stalk Synthetic. p. 83-104. In: J.P. Gustafson (ed.) Stadler genetics symposium, 15th, Columbia, MO

Hartl, D. L., and A.G. Clark. 1997. Principles of population genetics. 3rd ed. Sinauer, Massachusetts.

Hedrick, P.W. 2000. Genetics of populations. 2nd ed. Jones & Bartlett, Massachusetts.

Helms, T.C., A.R. Hallauer, and O.S. Smith. 1989. Genetic drift and selection evaluated from recurrent selection programs in maize. Crop Sci. 29: 602- 607.

Heredia-Diaz, O., A. Alsirt, L.L. Darrah, and E.H. Coe. 1996. Allelic frequency changes in the MoSCSSS maize synthetic in response to bi-directional recurrent selection for rind penetrometer resistance. Maydica 41: 65-76.

Holthaus, J.H., and K.R. Lamkey. 1995. Populations means and genetic variances in selected and unselected Iowa Stiff Stalk Synthetic maize populations. Crop Sci. 35: 1581-1589.

Kimura, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28: 882-901.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416-423.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1999. Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. 99: 1166-1178.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 2000. Hardy- Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45: 243-255.

Lamkey, K.R. 1992. Fifty years of recurrent selection in the Iowa Stiff Stalk Synthetic maize populations. Maydica 37: 19-28.

Lewis, P. O., and Zaykin, D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16b). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html

Messmer, M.M., A.E. Melchinger, M. Lee, W.L. Woodman, E.A. Lee, and Lamkey, K.R. 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize populations: comparison of allozyme and RFLP data. Theor. Appl. Genet 83: 97-107.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Nei, M., and S. Kumar. 2000. Molecular evolution and phylogenetics. Oxford University Press,New York.

Neuhausen, S.L. 1989. A survey of Iowa Stiff Stalk Synthetic parents, derived inbreds and BSSS(HT)C5 using RFLP analysis. Maize Genet. Coop. Newsl. 63: 110-111.

Popi, J., L. Rajnpreht., L.W. Kannenberg, and K.P. Pauls. 2000. Random amplified polymorphic DNA-based evaluation of diversity in the hierarchical, open-ended population enrichment maize breeding system. Crop Sci. 40: 619-625.

Russell, W.A. 1979. Registration of maize germplasm (Reg. No. GP 78 and GP 79). Crop Sci. 19: 565.

Smith, O.S. 1983. Evaluation of recurrent selection in BSSS, BSCB1, and BS13 maize populations. Crop Sci. 23: 35-40.

Sprague, G.F. 1946. Early testing of inbred lines of corn. J. Am. Soc. Agron. 38; 108-117.

Troyer, F. 2000. Temperate corn - background, behavior and breeding. p. 393 - 463. In Hallauer, A.R. (ed). Specialty corns. 2nd ed. CRC Press Inc., Boca Raton, USA.

Weir, B.S. 1996. Genetic data analysis II. 2nd ed. Sinauer, Massachusetts.

Wright, S. 1951. The genetical structure of populations. Ann. Eugen. 15: 323-354.

Table 1. RFLP loci evaluated in three BSSSP, BS13(S)C0, and BS13(S)C7 maize populations †.

agrr115	bnl07.49	phi10005	umc042b	umc095
asg008	bnl07.71	phi10016	umc043	umc097
asg024	bnl08.32	phi10017	umc046	umc107
asg045	bnl08.33	phi20725	umc052	umc108
asg062	bnl08.35	phi20854	umc054	umc110
bnl03.03	bnl09.11	umc004	umc055a	umc113
bnl03.04	bnl09.44	umc005	umc055b	umc120
bnl03.06	bnl12.09	umc007	umc057	umc121
bnl05.09	bnl12.30	umc015	umc059	umc128
bnl05.10	bnl13.05a	umc019	umc060	umc131
bnl05.14a	bnl13.05b	umc021	umc061	umc134
bnl05.14b	bnl15.07	umc026	umc062	umc137
bnl05.46	bnl15.40	umc027	umc065a	umc140
bnl05.47	csu061	umc030	umc065b	umc147
bnl05.62	csu147	umc031	umc076	umc152
bnl05.67	csu164	umc032	umc080	umc153
bnl05.71	ksu005	umc034	umc081	umc156
bnl06.06	npi097	umc035	umc084	umc157
bnl06.25	npi268	umc036	umc085	umc158a
bnl06.32	npi414	umc038	umc089	umc158b
bnl07.26	phi06005	umc042a	umc090	umc159

† Restriction enzyme HindIII was used in combination with the majority of the probes while EcoR1 was used in combination with the following probes: agrr115, asg045, asg062, bnl05.62, bnl12.30, csu061,csu164, ksu005, npi268, phi06005, phi10005, phi20725, umc005, umc021, umc055, umc060, umc076, umc121, and umc140.

Table 2. Statistics related to number and frequency of alleles estimated in BSSSP, BS13(S)C0, and BS13(S)C7 maize populations based on a sample of 105 RFLP loci.

	BSSSP	BS13(S)C0	BS13(S)C7
Proportion of polymorphic loci (<0.95)	0.98	0.79	0.66
Number of alleles	439	251	231
Mean allele frequencies	0.24 ± 0.01	0.42 ± 0.02	0.46 ± 0.02
Mean number of alleles per locus	4.18 ± 0.17	2.39 ± 0.09	2.20 ± 0.08
Number unique alleles	174	5	7

Table 3. F_{IS} coefficient and estimated and predicted values for mean expected heterozygosity (H_e) in BSSSP, BS13(S)C0, and BS13(S)C7 maize populations based on a sample of 105 loci.

Cycles	H_e † (estimated)	H_e ‡ (predicted)	F_{IS} coefficient §
BSSSP	0.58 ± 0.02		0.980 (0.967, 0.991)
BS13(S)C0	0.34 ± 0.02	0.41 ± 0.01	0.013 (-0.014, 0.039)
BS13(S)C7	0.26 ± 0.02	0.27 ± 0.02	0.018 (- 0.004, 0.045)

† H_e means average expected heterozygosity.

‡ Predicted by the equation $He_t = He_0 [1- (1/2Ne)]^t$ (see Materials and Methods)

§ Mean and (95% confidence interval) for F_{IS}, respectively.

Table 4. The number of loci for which pairs of populations shared or did not share the same most common allele (MCA). The frequency of the MCA is given for each population in a comparison †.

	BSSSP vs. C0	BSSSP vs. C7	C0 vs. C7
Coincident loci			
Number	56	57	78
MCA frequency in the earliest generation	0.61 ± 0.027	0.60 ± 0.026	0.80 ± 0.026
MCA frequency in the latest generation	0.80 ± 0.027	0.85 ± 0.024	0.86 ± 0.019
Non coincident loci			
Number	49	48	27
MCA frequency in the earliest generation	0.50 ± 0.022	0.52 ± 0.025	0.60 ± 0.026
MCA frequency in the latest generation	0.69 ± 0.023	0.77 ± 0.027	0.66 ± 0.031

† C0 and C7 means BS13(S)C0 and BS13(S)C7 populations, respectively.

Table 5. Correlation among three measures of genetic diversity (most common allele, number of alleles, and expected heterozygosity) estimated in three maize populations based on a sample of 105 RFLP loci †.

Comparison ‡	BSSSP	BS13(S)C0	BS13(S)C7
MCA vs. H_e	-0.96 ± 0.008	-0.94 ± 0.024	-0.96 ± 0.014
NA vs. H_e	0.78 ± 0.027	0.76 ± 0.039	0.70 ± 0.049
MCA vs. NA	-0.72 ± 0.041	-0.66 ± 0.064	-0.66 ± 0.058

† Standard errors estimated by jackknife.

‡ MCA, NA, and H_e means most common allele, number of alleles, and expected heterozygosity, respectively.

Table 6. F_{IS} coefficient and expected and observed heterozygosity for RFLP loci with significant deviations from Hardy-Weinberg proportions in BS13(S)C0 and BS13(S)C7 maize populations.

RFLP locus	Significance H-W test	F_{IS}	H_e	H_o
BS13(S)C0				
agrr115	***	0.106	0.671	0.600
asg062	*	-0.225	0.487	0.596
bnl03.03	**	-0.002	0.655	0.656
bnl08.33	**	0.520	0.115	0.056
csu061	***	0.375	0.545	0.341
phi06005	*	0.191	0.617	0.500
phi20854	*	0.112	0.325	0.289
umc057	***	0.379	0.572	0.356
umc095	*	-0.027	0.638	0.656
umc137	**	0.212	0.310	0.244
BS13(S)C7				
asg062	***	0.077	0.478	0.441
bnl05.09	*	0.223	0.479	0.373
bnl13.05a	*	0.664	0.029	0.010
umc007	**	0.303	0.528	0.489
umc065b	**	0.273	0.499	0.363
umc085	*	0.141	0.315	0.322
umc158a	*	0.299	0.154	0.108

*,**,*** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively.

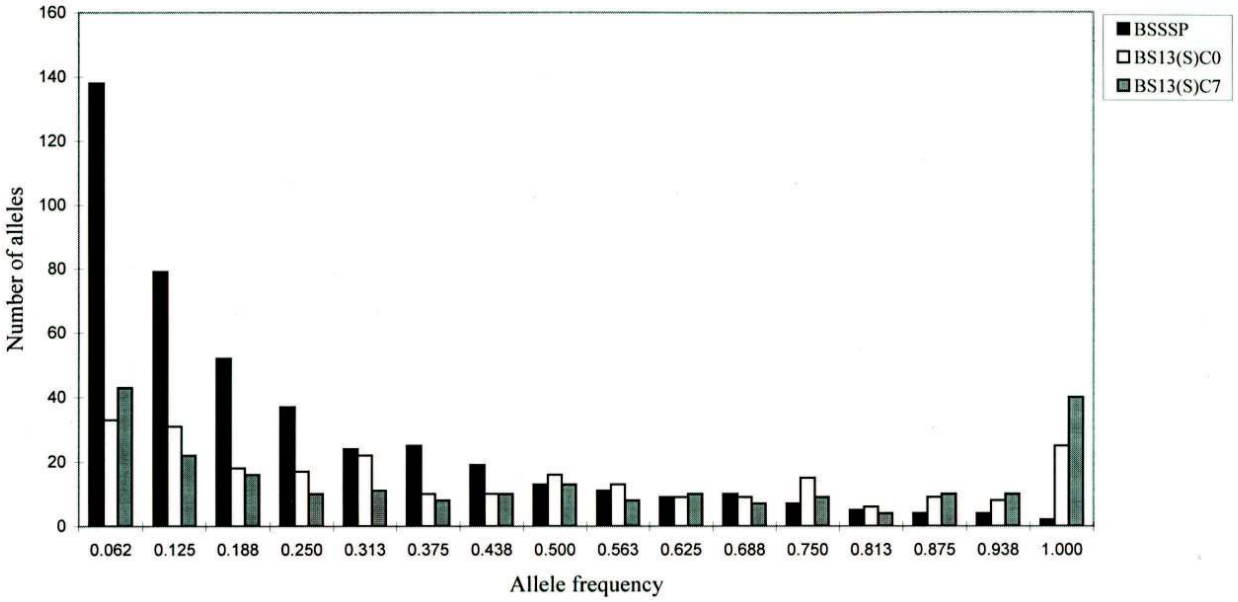

Figure 1. RFLP allele frequency distribution in BSSSP, BS13(S)C0, and BS13(S)C7 maize populations. Values in x-axis are upper boundaries for the class intervals.

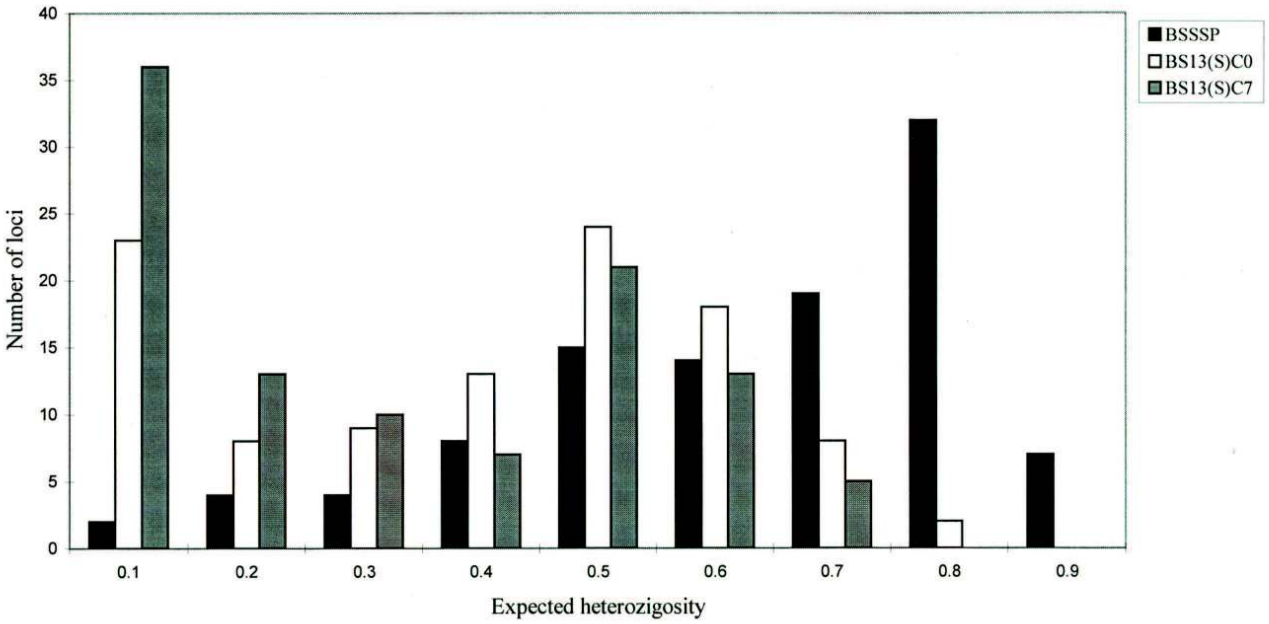

Figure 2. Distribution of expected heterozygosity by RFLP locus in BSSSP, BS13(S)C0, and BS13(S)C7 maize populations.

Values in x-axis are upper boundaries for the class intervals.

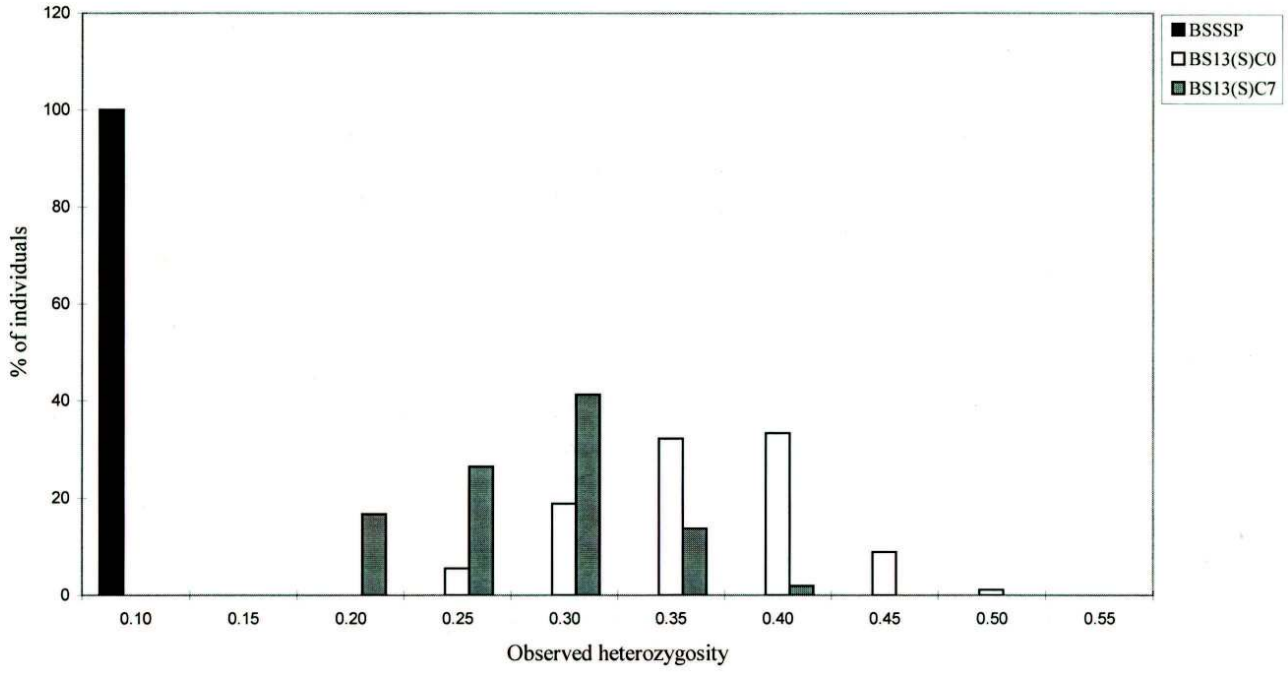

Figure 3. Distribution of observed heterozygosity by individual in BSSSP, BS13(S)C0, and BS13(S)C7 populations. Values in x-axis are upper boundaries for the class intervals.

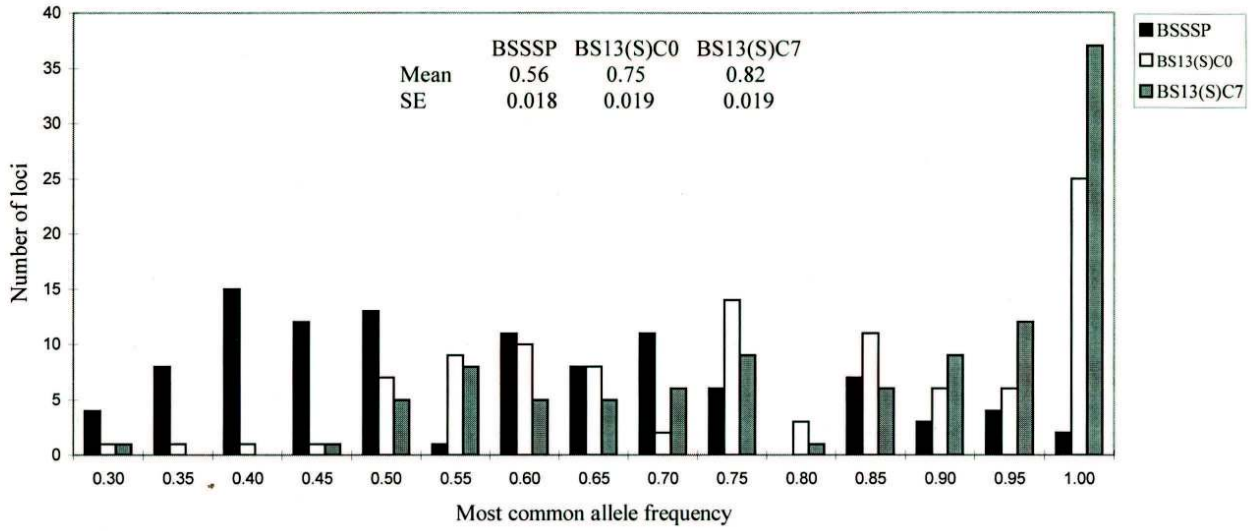

Figure 4. Distribution of the most common allele in BSSSP, BS13(S)C0, and BS13(S)C7 populations. The most common allele per locus was independently identified for each population. Values in x-axis are upper boundaries for the class intervals.

MOLECULAR MARKER DIVERSITY FOLLOWING HALF-SIB AND S2 SELECTION IN THE IOWA STIFF STALK SYNTHETIC MAIZE (*Zea mays L.*) POPULATION

A paper to be submitted to Crop Science

Paulo Evaristo de Oliveira Guimarães and Kendall Raye Lamkey

Abstract

The purpose of this study was to evaluate the genetic diversity among cycles of selection in the Iowa Stiff Stalk Synthetic maize population, which has been undergoing recurrent selection (RS) for more than 60 years. Three generations were used to evaluate seven cycles of half-sib selection followed by seven cycles of S_2 selection. BSSSP (16 progenitor lines), BS13(S)C0 (90 individuals from C7 of HS selection and C0 from S_2 selection), and BS13(S)C7 (102 individuals from the 7^{th} S_2 cycle) were genotyped with 105 RFLP loci. Principal Component Analysis and Nei's genetic distance (NGD) revealed C0 and C7 were close (NGD = 0.08) and apart from P (NGD = 0.26 and 0.33, respectively), indicating that the greatest changes occurred during HS selection. We also observed a considerable loss of genetic variation over the cycles of selection. Progenitor line Illinois Hy had a lower NGD to C0 and C7 and five of its unique alleles had significant increases in frequencies from BSSSP to C0 and C7. Hybrid Hy x LE 23 showed the lowest NGD to both, the C0 and C7 populations. NGD among progenitor lines was not a good predictor of single-crosses performances estimated in a previous study. Changes in allele frequencies for about 30% of the loci during the 14 cycles of selection cannot be explained by genetic drift alone. Future

studies could identify if there are loci linked with QTLs in this RS program via molecular and phenotypic evaluation of progenies representing different cycles of selection.

Introduction

History and evolution can be defined as "a branch of knowledge that records and explains past events" and "change in the gene pool of a population from generation to generation by such processes as mutation, natural selection, and genetic drift", respectively (Webster, 1996). Genetic gains, heritabilities, and components of variance need to be periodically estimated to provide information about the evolutionary history of populations under long-term recurrent selection (RS) programs. This knowledge can lead to modifications in selection procedures that will increase the probability of sustaining responses for the traits of interest in many generations.

Hallauer and Miranda (1988) compared a large number of maize breeding programs based on phenotypic data. They reported that response to selection can be affected in a variable way by factors such as: population, environment, breeding method, number of progenies, experimental design, cycles of selection, trait, number of traits, use of selection index, etc. Thus, due to this context dependency, selection response for quantitative traits is complex and requires more studies to understand changes in genetic variation of populations undergoing selection.

In addition to phenotypic data, allozyme (Brown, 1971; Brown and Allard, 1971; Stuber and Moll, 1972; Stuber et al., 1980; Kahler, 1983; Revilla et al., 1997) and molecular marker data (Heredia-Diaz et al., 1996; Labate et al., 1997, 1999, and 2000; Popi et al., 2000) can be used to monitor maize populations undergoing RS. Brown, 1971, Brown and Allard (1971), and Revilla et al. (1997) reported changes in allele frequencies that could be attributed to

genetic drift whereas Stuber et al. (1980), Heredia-Diaz et al. (1996), and Labate et al. (1999) found variation at some loci greater than what could be explained by genetic drift alone. Labate at al. (2000) and Popi et al. (2000) used principal component analysis (PCA) based on allele frequencies to evaluate reciprocal selection programs in maize. In both studies it was possible to distinguish cycles of selection based on this technique, revealing that PCA can be an alternative to evaluate genetic variation between and within cycles of selection. Labate et al. (1997) observed that reciprocal recurrent selection increased genetic distances within and between BSSS(R) and BSCB1(R) populations, whereas Revilla et al. (1997) did not find evidence of changes in genetic distances over cycles of selection in two synthetics undergoing intrapopulation improvement. One can observe that results based on allozyme and molecular data are, as expected, also variable and dependent on diverse factors such as: population, breeding method, effective population size, number of cycles, number of marker loci, type of marker used, etc.

The greater the number of genes involved in a trait, the greater the probability that this trait will show a large proportion of genes with low selective values (Comstock, 1978; Falconer and Mackay, 1996; Hartl and Clark, 1997). For quantitative traits, such as grain yield in maize, a large proportion of the genes may be effectively neutral. Neutral genes are more subject to random genetic drift, i.e., with changes in allele frequency less subject to selection and more dependent on initial allele frequency and effective population size (Crow and Kimura, 1970; Hedrick, 2000). Following this rationale, one can infer that changes among cycles of selection measured on neutral molecular markers can give some indication of magnitudes of changes in effectively neutral genes for the quantitative traits of interest because in both cases genetic drift is expected to be the major component of genetic

variation, especially if these markers and genes are not linked with others with strong selection effects.

Iowa Stiff Stalk Synthetic (BSSS) has been undergoing RS for more than 60 years. Although the breeding history of this maize population has been well recorded based on phenotypic data (Helms et al., 1989; Lamkey, 1992; Holthaus and Lamkey, 1995) there has been little evaluation of this program based on molecular data. The objective of our research was to evaluate the impacts of RS over the cycles of selection in BSSS population based on RFLP data driven by the following questions: (i) How genetically diverse are the cycles of selection based on genetic distances and principal component analysis? (ii) What was the genetic contribution of the progenitor lines to the cycles of selection? (iii) Are the genetic distances among progenitor lines good predictors of single-cross performance? (iv) Were there RFLP loci with changes in allele frequencies that were greater than could be explained by genetic drift alone?

Material and Methods

Genetic Material

Iowa Stiff Stalk Synthetic (BSSS) was formed by intermating 16 inbred lines previously selected for stalk quality (Sprague, 1946). In this study a collection of progenitor lines, named BSSSP or just P herein, was used to represent the original synthetic. This collection of progenitor lines included 14 of the 16 original lines plus the progenitor lines (Fe and IndB2) of line F1B1 (one of the missing lines). Neither line CI617 nor its progenitor lines was available to represent it. Messmer et al. (1991) reported that 10 of these lines were selected from the Reid Yellow Dent (RYD) population and six were selected from other populations (Table 1). Seven cycles of HS selection followed by nine cycles of S_2 selection for grain

yield and other traits has been conducted in this intrapopulation RS program. During the HS selection period the double cross hybrid IA 13 was used as tester, the harmonic mean of the number of selected progenies was 10 and the intensity of selection averaged 9.7% (Holthaus and Lamkey, 1995). During the six cycles of S_2 selection the harmonic mean of the number of selected progenies was 15 and the intensity of selection averaged 15.6% (Holthaus and Lamkey, 1995). A sample of the seventh HS and original S_2 cycle, named BS13(S)C0 or just C0 herein, and a sample of the seventh S_2 cycle, named BS13(S)C7 or just C7 herein, were used to evaluate this intrapopulation selection program. Syn-2 generations from BS13(S)C0 and BS13(S)C7 populations were obtained by random mating about 350 individuals from Syn-1 generations. For this study, 90 and 102 individuals from Syn-2 generations were used to represent BS13(S)C0 and BS13(S)C7 populations, respectively.

RFLP Analysis

Sampling of populations followed the approach used by Labate et al. (1997). Progenitor lines and the populations were grown in the same environment. Mature leaf tissue from randomly chosen individuals was collected, freeze-dried, ground, and stored at -20^0C. DNA isolation and RFLP analysis were performed at Biogenetic Service INC that also provided an initial summarization of the observed fragment data. DNA was isolated using a CTAB extraction method. Techniques for DNA isolation and other RFLP methods are found in Ausubel (1994) with slight modifications by Biogenetic Services, Inc. (unpublished). DNA was digested using two restriction enzymes, Hind III and EcoR1, depending on the probe employed. Gel electrophoresis was conducted with a TAE buffer. Southern blots were made on nylon membranes and the resulting complex of fragments was subject to hybridization techniques (Budowle and Baechtel, 1990). After washing the membranes to remove the

unhybridized probe solution, the membranes were placed on Kodak AR X-ray film and exposed for up to 14 days depending on the optimal exposure time. One hundred probe-enzyme combinations were used. Eighty-one and 19 of these combinations were made with RFLPs obtained by using restriction enzymes Hind III and EcoR1, respectively. The RFLP probes are distributed on all 10 maize chromosomes. Information about origin, sequence, location, and references of these probes can be obtained at Maize Genetic Database (http:// www.agron.missouri.edu). Molecular weight of each band was assigned based on Lambda markers. Each alternative band was considered as one allele. Some alleles, however, were associated with more that one band. Probe BNL05.24 was discarded because the banding patterns were too complicated to be interpreted. Six probes (bnl05.14, bnl13.05, umc042, umc055, umc 065, and umc158) were interpreted as mapping to two loci each. A total of 105 RFLP loci were detected (Table 2). Data were recorded in binary code and stored in a database developed by Frisch (M. Frisch, personal communication, 2000).

Principal Components Analysis

Graphical discrimination among all 208 individuals sampled was determined based on binary data of all 456 RFLP alleles observed. A correlation matrix was constructed and principal components were estimated by using the program SAS Analyst (SAS institute, 1999). According to Melchinger et al. (1991), principal components based on the correlation matrix are more efficient for discriminating among individuals than those based on the covariance matrix, because the correlation approach gives less weight of discrimination to alleles with higher variance across the individuals evaluated, i.e., alleles with intermediate frequencies, and more weight to alleles that show lower variance, i.e., alleles with extreme frequencies. The scores of the first three principal components, estimated for each individual

from all three populations, were used to construct a three-dimensional graphic by using the program SPLUS (MathSoft, 1997). Another 3-D graphic was constructed including just individuals from C0 and C7 populations.

Neutrality Tests

Waples (1989) developed a test for determining whether temporal changes in allele frequency deviate from the expected by genetic drift. Let x_0 and x_t be the frequencies of an allele sampled at different generations. The quantity $(x_0-x_t)^2 / V(x_0-x_t)$ follows a X^2 distribution and can be used to test a null hypothesis that genetic drift and sampling error can explain temporal changes in allele frequencies. $V(x_0-x_t)$ is composed of genetic and statistical sampling components. Sampling Plan II of Waples (1989) was used, because the sampled plants were different from the ones that participated in the reproduction. Let Ne be the effective population size, estimated as the harmonic mean of the number selected progenies intermated in successive generations, "t' the number of generations, and S_n the sample size at generation N. These measures are necessary to estimate $V(x_0-x_t)$. TEMPTEST software (R. Waples, personal communication) was used for performing Waples' test considering all 14 cycles of selection (Ne = 12, t = 14, S_0 = 16, S_7 = 90, and S_{14} = 102). TEMPTEST can evaluate up to three alleles per locus at several sampled generations. A program was developed (K. R. Lamkey, personal communication) using SAS-IML language (SAS Institute, 1990) to evaluate multiple alleles per locus in two sampled generations. This software was used for performing Waples' test for the 7 cycles of HS selection (Ne = 10, t = 7, S_0 = 16, and S_7 = 90) and also for the 7 cycles of S_2 selection (Ne = 15, t = 7, S_7 = 90, and S_{14} = 102).

Genetic Distances Among Populations

Genetic Data Analysis software (Lewis and Zaykin, 2001) was used for computing the traditional Nei (1978) unbiased estimate of standard genetic distance (NGD) among populations. This genetic distance is estimated as NGD = -ln*I* , where *I* is genetic identity. When allele frequencies are identical between two populations, *I* equals one and NGD equals zero, and when an allele is fixed in one population and absent in the other population, *I* equals zero and NGD is ∞ (Nei 1987, page 221). Genetic distances that included C0 and C7 cycles were estimated as the average of NGD measured on an individual basis, i.e., 90 individuals in CO and 102 in C7. By using this approach, it was possible to estimate standard errors and confidence intervals for NGD by using resampling techniques over individuals. Standard errors were estimated via jackknifing and 95% confidence intervals (2.5 and 97.5 bias-corrected and adjusted, BCa, percentiles) were obtained based on 1000 bootstrap samples (Efron and Tibshirani, 1986; Mathsoft, 1997).

Relative Divergence Of Progenitor Lines To Cycles Of Selection

A data matrix was constructed with all 120 NGD between the 16 progenitor lines. The genetic divergence between a progenitor line and BSSSP was estimated as the average Nei (1978) genetic distance of all 15 NGD between lines that included that progenitor line. The genetic divergence of each progenitor line to BSS13(S)C0 was estimated as the average NGD of each progenitor line with all 90 C0 individuals. The genetic divergence of each progenitor line to BS13(S)C7 was estimated as the average NGD of each progenitor line with all 102 C7 individuals. The standard errors for NGD were also estimated via jackknifing. All the correlations between average NGD of the 16 progenitor lines to P, C0, and C7 were

computed to determine if there was an association of genetic divergence of the progenitor lines estimated in different cycles of selection.

Based on measurements in C0 individuals, correlation coefficients were estimated for genetic distances between each progenitor line and C0 versus genetic distances between P and CO. Based on measurements in C7 individuals, correlation coefficients were estimated for genetic distances between each progenitor line and C7 versus genetic distances between P and C7. This approach was used to determine if there was association between the genetic divergence of each progenitor line and the genetic divergence of BSSSP to the advanced cycles of selections.

A data matrix was constructed with the genotypes of all 120 F1 hybrids among the 16 progenitor lines. The genotypes of progenitor lines were used to predict the genotype of the corresponding F1 obtained by crossing two of the progenitor lines. Driven by the question " Which F1 hybrids show the lowest divergence to the advanced cycles of selection?", we estimated NGD of each F1 hybrid to C0 (average of 90 individuals) and C7 (average of 102 individuals).

Stucker and Hallauer (1992) evaluated the combining ability for grain yield and other agronomic traits for the 16 BSSS progenitor lines in field trials conducted for 2 years at three Iowa locations. Each progenitor line was crossed with 12 other progenitor lines in six sets of 4 x 4 Design II crosses, producing a total of 96 crosses. To determine if the genetic distances among progenitor lines are good predictors of grain yield single-crosses performance, the 96 NGD for the F1's were estimated and correlated with the 96 estimates of F1 grain yield obtained by Stucker and Hallauer (1992). Also, NGD of the 96 F1 hybrids to C0 and C7 cycles of selection were correlated with the 96 estimates of F1 grain yield obtained by

Stucker and Hallauer (1992). This approach was used to determine if F1 grain yield performance was associated with the genetic divergence of the F1 hybrids to the advanced cycles of selection.

The 120 hybrids were divided into 16 groups containing the 15 hybrids that share a common parent. NGD was estimated between each group of F1 hybrids and P (16 lines as a group), C0 (average of 90 individuals), and C7 (average of 102 individuals) to answer the question " Which group of F1 hybrids had the lowest divergence to the cycles of selection?" Correlations (N=16) between average NGD of the 16 groups of F1 hybrids and P, C0, and C7 were computed to determine if there was an association of genetic divergence of the F1 hybrids estimated in different cycles of selection. Also, NGD between the 16 groups of F1 hybrids and P, C0, and C7 cycles of selection were correlated with the 16 estimates of general combining ability effects for grain yield obtained by Stucker and Hallauer (1992). This approach was used to determine if general combining ability effects for grain yield were associated with the genetic divergence of F1 hybrids groups to the cycles of selection.

Unique alleles are those alleles that trace exclusively to a single progenitor line. The change in frequency of unique alleles over cycles of selection provided additional information about the contribution of a progenitor line to populations developed via selection.

Results and Discussion

Principal Components Analysis (PCA)

The first three principal components explained 12.2, 6.9, and 4.8% of the variation among all 208 individuals at all 105 loci (Fig. 1). There was no overlap of individuals from different populations. The C0 and C7 populations were closely located and showed little

within population variability. The P group showed the highest variation and was separated from C0 and C7. The group of progenitor lines that were derived from RYD showed less variation than the non- RYD lines (Fig. 1). The non-RYD lines had more unique alleles than the RYD lines, indicating that the non-RYD group made a significant contribution to the variation observed in this synthetic (Table 1). Messmer et al. (1991) genotyped the BSSSP lines with allozymes and RFLPs. Their results also indicated that non-RYD lines played an important role in the variation of the BSSSP population. The lines A3G-3-1-3, Ill. 12E, and OH3167B were positioned the farthest from other lines (Fig. 1). It is interesting to note that the origin of lines A3G-3-1-3 and Ill. 12E is unknown and that they had the highest number of unique alleles (Table 1). Plotting the first three principal components for just BS13(S)C0 and BS13(S)C7 gives a better visualization of the distribution of the individuals from these populations (Fig. 2). The C0 and C7 populations were located closely to one another, but there was no overlap of their individuals and the variability was lower within C7 than C0. Consistent with the graphical representation, the standard errors of the three principal components were higher in P than CO and C7, higher in non-RYD than RYD group, and in close agreement between C0 and C7 (Table 3). In comparison to RYD, the non-RYD group was more distant from CO in the first and third principal components, and less distant from C7 in the second and third components. RYD and non-RYD were more distant in the second component while C0 and C7 were more distant in the third component.

PCA results indicated a clear separation among individuals of different cycles of selection. These results are consistent with those obtained by Labate et al. (2000). They performed PCA to evaluate a reciprocal recurrent selection program between BSSS and

BSCB1 populations and observed that after 12 cycles of selection there was a clear distinction of individuals belonging to the different populations and cycles of selection.

Our PCA results indicate that the greatest changes occurred during HS selection since there was considerable distance between C0 and P and the within variability was reduced drastically. The HS selection program was characterized by a low effective population size (harmonic mean of 10), and a high selection intensity (9.7%) (Holthaus and Lamkey, 1995). Thus, the magnitude of changes during HS selection may be explained by genetic drift and selection from the base population, BSSSP, which contained many unique alleles and abundant variation. The S_2 selection program was characterized by a low effective population size (harmonic mean of 15), and a selection intensity of 15.6% (Holthaus and Lamkey, 1995). The reduced variation in BS13(S)C0 may explain why genetic drift and selection during S_2 selection caused less divergence of C7 from C0 than what occurred between C0 and P.

Dubreil and Charcosset (1998) suggested that a mixture of individuals from different populations could explain two clusters of individuals they observed in BS13(S)C4. We did not find distinct clusters within the C0 or C7 populations (Fig. 2). Labate et al. (2000) observed a few individuals in the BSSS(R)C12 population that had a high frequency of rare alleles. They also suggested that mixing of individuals from different populations could be a reasonable explanation for their results by suspecting that the individuals in question came accidentally from the BSSS(R)C0 population. Dubreil and Charcosset (1998, 1999) also found that BS13(S)C4 was distinct from other US populations. Probably, the long-term recurrent selection in BS13(S) contributed to its divergence from other populations. Further

studies could address this question by evaluating the genetic divergence of different cycles of BSSS from other maize populations.

Tests For Changes In Allele Frequency

Fourteen loci had changes in allele frequencies that were greater than what would be expected from genetic drift and sampling error alone during HS selection (Table 4). Twenty-three loci showed significant changes in allele frequencies after seven cycles of S2 selection (Table 4). When all the 14 cycles of selection were considered together, the null hypothesis that genetic drift was responsible for the observed change in allelic frequency was rejected at ~ 30% of the loci. This evaluation involved the sampling of three generations (P, C0, and C7) and provided a more powerful test of changes in allele frequencies than considering either the HS period or the S_2 selection period individually. Thus, there is evidence that a considerable fraction of the loci have changes in allele frequencies that cannot be explained by random genetic drift in this recurrent selection program.

Fourteen cycles of selection can be considered as a short period of time for mutation be a major source of allele variation and the process of selection was made with considerable caution to avoid migration. Also, if either or both of these factors were important, then we could expect to observe a considerable proportion of unique alleles in the selected populations. However, in a previous study (to be published) we found just a small proportion of unique alleles in C0 and C7 populations. Hence, selection, either directly or by hitchhiking, probably was the major force changing allele frequencies at those loci where the null hypothesis of genetic drift alone was rejected.

Labate et al. (1999) evaluated changes in allele frequencies of 82 RFLP loci in two populations undergoing reciprocal recurrent selection, BSSS(R) and BSCB1(R). They

rejected the null hypothesis of genetic drift for about 17% of the loci in each population. Six RFLP loci (bnl09.44, umc030, umc034, umc054, umc108, and umc128) in BSSS showed significant deviations from genetic drift in our study and in Labate's study. Four RFLP loci (bnl06.32, bnl07.71, bnl12.09, and umc110) showed significant deviations from genetic drift and were in common to the BS13 and BSCB1 populations in our and Labate's study, respectively. These results lend additional support to the hypothesis that these 10 loci showed allele frequencies changes greater than could be expected in a pure genetic drift model. It would be interesting to use these 10 nonneutral loci on the response for the traits undergoing selection in a marker assisted selection program based on the original BSSS population. Future studies could address the question about which loci are linked with QTLs for the traits of interest by evaluating progenies from different cycles of selection based on molecular and phenotypic data.

Genetic Distances Between Cycles Of Selection

As expected, one of the consequences of long-term selection was genetic divergence between the cycles of selection. Nei's (1978) genetic distance (NGD) between P and C0 was greater than between C0 and C7 (Table 5). These results are in agreement with those observed for PCA, also suggesting that the major changes in genetic divergence occurred during the HS selection period. Labate et al. (1997) also reported an increase in divergence between cycles of reciprocal recurrent selection. They observed that NGD between the progenitor and 12th cycles of selection were 0.33 and 0.26 in populations BSSS(R) and BSCB1(R), respectively. RYD and non-RYD progenitor showed similar NGD values to C0 and also to C7 (Table 5), suggesting that both groups had comparable contribution to the variation of the advanced cycles of selection.

According to Hedrick (1999, 2000), the standard genetic distance between an ancestral and a descendent population can be predicted as:

$NGD = - 0.5 \ln [(1- H_0) / (1 - H_0 (1 - 1 / 2Ne)^t)]$,

where H_0, Ne, and t stands for the initial expected heterozygosity, effective population size, and number of generations, respectively. In our study, Ne was assumed as 10 and 15 during the HS and S_2 selection periods, respectively. In a previous report (to be published), H_0 was estimated as 0.58 and 0.34 for generations P and C0, respectively. Given these estimates, NGD was predicted as 0.17 between P and C0 and 0.05 between C0 and C7. The predicted value was in close agreement with the observed for the distance between C0 and C7, but lower than the observed for the distance between P and C0. Labate et al. (1999) reported that Ne estimated from nonneutral loci was lower than the number of selected progenies in BSSS and BSCB1 maize populations after 12 cycles of reciprocal recurrent selection. The observed NGD between P and C0 is larger than expected, suggesting that the Ne during the HS period was lower than the harmonic mean of the number of selected progenies. This result can be considered as an additional indicator that selection acted directly or indirectly on some of the RFLP loci during the HS period. The same equation was used to predict NGD between C7 and C14 (a future 14^{th} cycle of S_2 selection). NGD between C7 and C14 was predicted as 0.03, assuming an average Ne = 20 and $H_0 = 0.26$. Thus, future cycles of selection are not predicted to generate much additional genetic divergence.

Relative Divergence Of Progenitor Lines To Cycles Of Selection

The greater the divergence of an individual from a population, the greater the genetic distance will be between them. There was large variation in estimates of NGD between the

individual progenitor lines and the C0 and C7 populations (Table 6). Lines Ind.AH83, Ind.B2, and Ind.461-3 consistently showed low NGD while lines A36-3-1-3,CI540, and I159 exhibited high NGD to P, C0, and C7 populations, respectively (Table 6). The correlations between estimates of for NGD computed from different cycles of selection were high (Table 10), i.e. in general, lines with higher or lower NGD to P and C0 also showed higher or lower NGD to C0 and C7, respectively. The Line Ill. Hy, however, showed a different pattern of genetic divergence, because it had one of the largest NGD to P, but had the lowest NGD to C7, and a lower than average NGD to C0 (Table 6). Moreover, line Hy was less distant to C7 than C0 while all other lines showed more divergence to C7 than C0.

F1 hybrids groups involving lines Ill. Hy and Ind.AH83 showed the lowest genetic distance to CO and the hybrids involving lines Ill. Hy and I159 had the lowest genetic distances to C7 (Table 7). The five F1 hybrids with the lowest NDG to C7 all involved line Ill Hy (Table 8). The hybrid Ill. Hy x LE 23 had the lowest NGD to both the C0 and C7 populations (Table 8). The results of Stucker and Hallauer (1992) showed that the single-cross Ill. Hy x LE 23 had a positive specific combining ability for yield and, based on the general combining ability (GCA) of Hy and LE 23 for other traits, one can infer that this single-cross probably had acceptable performance for root and stalk lodging and plant and ear height. Additional research would be needed to identify whether selection or genetic drift was responsible for Hy x LE23 having the lowest NGD to C0 and C7 populations.

To obtain information about the genetic divergence relationship between a line and P, correlations among genetic distances were estimated based on individuals of the C0 and C7 populations. The correlations ranged from 0.51 to 0.78, however there was one exception: the correlation of (NGD line Ill. Hy vs. C0) versus (NGD P vs. C0) was 0.36 (Table 9). One

explanation for this result and also for the low NGD between Hy and the C0 and C7 populations, is that alleles from Hy have been preferentially selected resulting in Hy making a greater contribution to the advanced cycles of selection than other progenitors. Line Hy also had the greatest frequency of unique alleles (0.55) that had been retained in C7 (Table 1), further indicating that Hy has made important contributions during selection. The six unique alleles from line Hy that were retained in C7 also had the highest average allele frequency in C7 and the second highest frequency in C0. Of the six remaining alleles, one was still rare in C7, whereas the frequencies of the other five increased from an average of 0.06 in P to 0.65 in C0 and 0.80 in C7. The changes in frequencies of these five alleles that traced exclusively to line Hy were higher than could be expected in a pure genetic drift model and their respective loci (bnl05.14a, bnl07.26, bnl07.71, bnl13.05a, and umc062) were classified as nonneutrals in Waples' test (Table 4), indicating that these five alleles may be shown selective advantage, directly or by hitchhiking, for trait(s) undergoing selection.

Previous studies have identified line Hy to have positive combining ability for grain yield and a low genetic divergence to later cycles of selection based on molecular markers. Line Hy had the fourth largest general combining ability effect for grain yield of the 16 progenitor lines (Stucker and Hallauer, 1992). The specific combining ability of the Kolkmeier and Lancaster populations improved when Hy was used as a tester in a recurrent selection program (Walejko and Russell, 1977). Neuhausen (1989) observed that Hy was one of the five progenitor lines that had unique alleles in a group of 3 (B73, B78, and B84) elite lines developed from HS selection. In comparison to other BSSS progenitor lines, line Hy showed the lowest genetic distance to elite line B84. However, the average genetic distance of this line to four elite BSSS lines (B14a, B37, B73, and B84) was similar to the average of all

progenitor BSSS lines (Messmer et al., 1991). Labate et al. (1997) reported that lines Hy and CI540 made significant contributions to the original and 12th cycles of reciprocal recurrent selection in the BSSS population. The relative contribution of progenitor lines to populations undergoing selection is a subject pertinent to breeding that could be the focus of future studies. One way to approach this problem would be to use molecular and phenotypic data to map QTLs associated with alleles that are unique to a progenitor line.

The correlations between estimates of NGD of different cycles of selection to the 16 progenitor lines were positive and significant (Table 10). These results are also consistent with those of Labate et al (1997) who reported high correlations for estimates of genetic contributions of progenitor lines to original and 12th cycles of selection in two maize populations. These results have implications to the development of populations because they indicate that progenitor lines more divergent from a given synthetic undergoing selection are also expected to be more divergent from the advanced cycles of selection. General combining ability for grain yield was negatively, but not significantly, correlated with the estimates of genetic distances of F1 hybrids groups to P, C0, and C7. Also F1 grain yield was negatively, but not significantly, correlated with the estimates of genetic distances between F1 hybrids to C0 and C7 (Table 10). However, NGD of F1 hybrids groups estimated in P and CO and also in C0 and C7 were positive and significant (Table 10). These results indicated that the divergence of genotypes to C0 and C7 were more associated with their genetic divergence to P and C0, respectively, than to their effects for grain yield. However, the results of correlation of genetic distances to cycles of selection and effects on grain yield need to be considered with caution for two reasons. First, Iowa Stiff Stalk Synthetic was formed by a series of eight single-crosses, four double-crosses, and one double-double cross (Labate et

al., 1999). Thus, although each progenitor line was equally represented, all 120 combinations among the 16 lines were never produced. Second, the estimates of general combining ability effects would be more accurate if they were based on 15 crosses for each line, instead of the 12 crosses used by Stucker and Hallauer (1992).

Nei's genetic distance among the 16 progenitor lines ranged from 0.11 (Ind.461-3 vs. I224) to 1.37 (A3G-3-1-3 vs. Ind.B2) (Table 11). This is in agreement with the PCA results and is another indicator that BSSS has considerable genetic variation. Messmer et al. (1991) estimated Rogers' genetic distance among the same group of lines based on 144 RFLP loci. They also found a substantial variation for genetic distances in the progenitor lines of Iowa Stiff Stalk Synthetic. The correlation between our estimates and Messmer's et al. (1991) estimates of genetic distances among the progenitor lines was 0.43 ($p < 0.001$), which indicates that genetic distances among a set of individuals is dependent on the sample of RFLP loci and the methodology used to estimate genetic distances.

The correlation between the F1 grain yield, estimated by Stucker and Hallauer (1992), and NGD among the progenitor lines was low (0.05, $p = 0.61$). We calculated the correlation between the Rogers' genetic distance, estimated by Messmer et al. (1991), and single-cross yield, estimated by Stucker and Hallauer (1992), as 0.17 ($p = 0.10$). Thus, genetic distances among these lines were not good predictors of their single-crosses yield performance. Other studies have reported correlations between F1 grain yield and genetic distances that were less than or equal to 0.5 (Lamkey et al. 1987, Lee et al. 1989; Godshalk et al. 1990; Melchinger et al. 1990a and 1990b; Dudley et al. 1991; Ajmone Marsan et al. 1998) whereas Smith et al. (1990) observed a coefficient of determination of 0.87 and Lanza et al. (1997) found two correlations higher than or equal to 0.7. Low heritability, weak dominance effects, low

proportion of markers linked to QTLs, and positive correlation for allele frequencies among progenitor lines for a given trait are some of the reasons raised by Bernardo (1992) for the low correlations between genetic distances and F1 performance.

Our results and those of other investigations indicated that molecular markers are useful tools to evaluate maize recurrent selection programs. More knowledge, however, is generated when molecular and phenotypic variation are evaluated jointly over the cycles of selection. If different programs of recurrent selection were evaluated based on phenotypic data and also with a standard set of marker loci, great advances could be obtained in the study of populations, breeding methods, traits undergoing selection, environments, and marker loci. This knowledge could lead to the design of more efficient maize recurrent selection programs based on phenotypic and marker assisted selection.

References

Ausubel, F (ed) 1994. Current protocols in molecular biology. John Wiley and Sons, Inc.

Ajmone Marsan, P., P. Castiglioni, F. Fusari, M. Kuiper, and M. Motto. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96: 219-227.

Bernardo, R. 1992. Relationship between single-cross performance and molecular marker heterozygosity. Theor. Appl. Genet. 83: 628-634.

Brown, A.H.D. 1971. Isozyme variation under selection in *Zea mays*. Nature 232: 570-571.

Brown, A.H.D. and R.W. Allard. 1971. Effect of reciprocal recurrent selection for yield on isozyme polymorphisms in maize (*Zea mays L.).* Crop Sci. 11: 888-893.

Budowle, B., and Baehtel, FS. 1990. Modifications to improve the effectiveness of restriction fragment polymorphism typing. Appl Theor Electrophoresis 1: 181-187

Comstock, R.E. 1978. Quantitative genetics in maize breeding. In Walden, D.B. (ed). Maize breeding and genetics. John Wiley & Sons, New York, New York.

Crow, J.F. and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Dubreil, P. and A. Charcosset. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor. Appl. Genet. 96: 577-587.

Dubreil, P. and A. Charcosset. 1999. Relationships among maize inbred lines and populations from European and North-American origins as estimated using RFLP markers. Theor. Appl. Genet. 99: 473-480.

Dudley, J.M., M.A. Saghai Maroof, and G.K. Rufener. 1991. Molecular markers and grouping of parents in a maize breeding program. Crop Sci. 31: 718-723.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard error, confidence intervals, and other measures of statistical accuracy. Statistical Science 1: 54-77.

Falconer, D.S. and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, England.

Godshalk, E.B., M. Lee, and K.R. Lamkey. 1990. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor. Appl Genet. 80:273-280.

Hallauer, A.R. and J.B Miranda Fo. 1988. Quantitative genetics in maize breeding. 2nd ed. Iowa State Press, Ames, Iowa.

Hartl, D.L. and A.G. Clark. 1997. Principles of population genetics. 3rd ed. Sinauer, Massachusetts.

Hedrick, P.W. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313-318.

Hedrick, P.W. 2000. Genetics of populations. 2nd ed. Jones & Bartlett, Massachusetts.

Helms, T.C., A.R. Hallauer, and O.S. Smith. 1989. Genetic drift and selection evaluated from recurrent selection programs in maize. Crop Sci. 29: 602- 607.

Heredia-Diaz, O., A. Alsirt, , L.L. Darrah, and E.H. Coe. 1996. Allelic frequency changes in the MoSCSSS maize synthetic in response to bi-directional recurrent selection for rind penetrometer resistance. Maydica 41: 65-76.

Holthaus, J.H. and K.R. Lamkey. 1995. Populations means and genetic variances in selected and unselected Iowa Stiff Stalk Synthetic maize populations. Crop Sci. 35: 1581-1589.

Kahler, A.L. 1983. Effect of half-sib and S1 selection for increased grain yield on allozyme polymorphism in maize. Crop Sci. 23: 572- 576.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman.1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416-423.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1999. Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. 99: 1166-1178.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 2000. Hardy- Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45: 243-255.

Lamkey, K.R. 1992. Fifty years of recurrent selection in the Iowa Stiff Stalk Synthetic maize populations. Maydica 37: 19-28.

Lamkey, K.R, A.R. Hallauer, and A.L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. Journal of Heredity 78: 231-234.

Lanza L. L. B., C. L. de Souza Jr., L. M. M. Ottoboni, M. L. C. Vieira, L. L. B., A. P. de Souza. 1997. Genetic distance of inbred lines and prediction of maize single-cross performance using RAPD markers. Theor. Appl. Genet. 94: 1023-1030.

Lee, M., E.B. Godshalk, K.R. Lamkey, and W.W. Woodman. 1989. Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci. 29: 1067-1071.

Lewis, P. O. and D. Zaykin. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16b). Free program distributed by the authors over the internet from http://lewis.eeb.uconn.edu/lewishome/software.html

MathSoft. 1997. S-PLUS 4.0 Release 2 for Windows. MathSoft, Massachusetts.

Messmer, M.M., A.E. Melchinger, M. Lee, W.L. Woodman, E.A. Lee, and K.R. Lamkey. 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize populations: comparison of allozyme and RFLP data. Theor. Appl. Genet. 83: 97-107.

Melchinger, A.E., M. Lee, , K.R. Lamkey, and W.L. Woodman. 1990a. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize hybrid. Crop Sci. 30: 1033-1040

Melchinger, A.E., M. Lee, K.R. Lamkey, A.R. Hallauer, and W.L. Woodman. 1990b. Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor. Appl. Genet. 80: 488-496.

Melchinger, A.E., M.M. Messmer, M. Lee, W.L. Woodman, E.A. Lee, and K.R. Lamkey. 1991. Diversity and relationships among U.S. maize inbreds revealed by restriction fragment length polymorphisms. Crop Sci. 31: 669-678.

Nei, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583:590.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Neuhausen, S.L. 1989. A survey of Iowa Stiff Stalk Synthetic parents, derived inbreds, and BSSS(HT)C5 using RFLP analysis. Maize Gen. Coop. Newsl. 63: 110-111.

Popi, J., L. Rajnpreht, L.W. Kannenberg, and K.P. Pauls. 2000. Random amplified polymorphic DNA-based evaluation of diversity in the hierarchical, open-ended population enrichment maize breeding system. Crop Sci. 40: 619-625.

Revilla, P., M.L. Vales, R.A. Malvar, and A. Ordas. 1997. Allozyme frequencies, heterozygosity and genetic distances following S_1 recurrent selection in two synthetic maize populations. Theor. Appl. Genet. 95: 1057-1061.

SAS institute. 1990. SAS/IML user's guide. Release 6.03 edition. SAS Institute Inc., Cary, North Carolina.

SAS institute. 1999. The Analyst Application, 1st ed. SAS Institute Inc., Cary, NC, USA

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis, and RFLPs. Theor. Appl. Genet. 80: 833-840.

Stuber, C.W., and R.H Moll. 1972. Frequency changes of isozymes alleles in a selection experiment for grain yield in maize (*Zea mays L.*). Crop Sci. 12: 337-340.

Stuber, C.W., R.H Moll, M.M.Goodman, H.E. Schaffer, and B.S. Weir. 1980. Allozyme frequency changes associated with selection for increased grain yield in maize (*Zea mays L.*). Genetics 95: 225-237.

Stucker, D.S. and A.R. Hallauer. 1992. Genetic variability as affected by selection in Iowa Stiff Stalk Synthetic maize. Journal of Heredity 83: 410-418.

Sprague, G.F. 1946. Early testing of inbred lines of corn. J. Am. Soc. Agron. 38; 108-117.

Waples, R.S. 1989. Temporal variation in allele frequencies: testing the right hypothesis. Evolution, 43: 1236-1251.

Walejko, R.N. and W.A. Russell. 1977. Evaluation of recurrent selection for specific combining ability in two open-pollinated maize cultivars. Crop Sci. 647-651.

Webster's talking dictionary / Thesaurus. 1996. Version 1.0b. Parsons Technologic Inc - Exceller Software Corp.

Table 1. Progenitor lines representing BSSSP population: origin, number of unique alleles, and number and average frequency of unique alleles that still remain in later generations.

Line	Origin†	Number of unique alleles ‡	Unique alleles in later generations §		
			Number	Average frequency	
				BS13(S)C0	BS13(S)C7
A3G-3-1-3	Unknown	30	6	0.03	0.02
CI540	Illinois Two Ear	10	4	0.14	0.23
Ill. Hy	Illinois High Yield	11	6	0.55	0.67
Ill.12E	Unknown	14	1	0.61	0.28
LE23	Illinois Low Ear	10	4	0.22	0.14
Oh3167B	Echelbeger Clarage	12	2	0.36	0.06
CI187-2	Reid Yellow Dent	6	3	0.35	0.25
I159	Reid Yellow Dent	4	1	0.00	0.01
I224	Reid Yellow Dent	0	0	-	-
Ind.461-3	Reid Yellow Dent	0	0	-	-
Ind.AH83	Reid Yellow Dent	3	0	-	-
Ind.B2	Reid Yellow Dent	9	3	0.11	0.17
Ind.Fe21073	Reid Yellow Dent	6	3	0.05	0.02
Ind.Tr9-1-1-6	Reid Yellow Dent	6	1	0.01	0.00
Os420	Reid Yellow Dent	10	4	0.12	0.13
WD546	Reid Yellow Dent	10	4	0.33	0.38

† Messmer et al. (1991)

‡ Alleles traced exclusively to a progenitor line.

§ Unique alleles from a progenitor line that were not lost in later generations.

Table 2. RFLP loci evaluated in BSSSP, BS13(S)C0, and BS13(S)C7 maize populations†

agrr115	bnl07.49	phi10005	umc042b	umc095
asg008	bnl07.71	phi10016	umc043	umc097
asg024	bnl08.32	phi10017	umc046	umc107
asg045	bnl08.33	phi20725	umc052	umc108
asg062	bnl08.35	phi20854	umc054	umc110
bnl03.03	bnl09.11	umc004	umc055a	umc113
bnl03.04	bnl09.44	umc005	umc055b	umc120
bnl03.06	bnl12.09	umc007	umc057	umc121
bnl05.09	bnl12.30	umc015	umc059	umc128
bnl05.10	bnl13.05a	umc019	umc060	umc131
bnl05.14a	bnl13.05b	umc021	umc061	umc134
bnl05.14b	bnl15.07	umc026	umc062	umc137
bnl05.46	bnl15.40	umc027	umc065a	umc140
bnl05.47	csu061	umc030	umc065b	umc147
bnl05.62	csu147	umc031	umc076	umc152
bnl05.67	csu164	umc032	umc080	umc153
bnl05.71	ksu005	umc034	umc081	umc156
bnl06.06	npi097	umc035	umc084	umc157
bnl06.25	npi268	umc036	umc085	umc158a
bnl06.32	npi414	umc038	umc089	umc158b
bnl07.26	phi06005	umc042a	umc090	umc159

† Restriction enzyme HindIII was used in combination with the majority of the probes while EcoR1 was used in combination with the following probes: agrr115, asg045, asg062, bnl05.62, bnl12.30, csu061, csu164, ksu005, npi268, phi06005, phi10005, phi20725, umc005, umc021, umc055, umc060, umc076, umc121, and umc140.

Table 3. Means and standard errors (SE) for the scores of the first three principal components (P1, P2, and P3), in three cycles of BSSS population and in two groups of progenitor lines (RYD and non-RYD) from BSSSP population.

Population	Scores					
	P1		P2		P3	
	Mean	SE	Mean	SE	Mean	SE
BSSSP	3.22	1.30	-1.05	3.52	-0.24	1.75
- Non-RYD	3.88	1.99	0.57	5.61	-0.64	2.96
- RYD	2.82	0.35	-2.02	0.54	-0.01	0.30
BS13(S)C0	-0.17	0.10	-0.04	0.07	0.97	0.23
BS13(S)C7	-0.36	0.07	0.20	0.05	-0.82	0.17

Table 4. Loci that showed changes in allele frequencies significantly higher that could be expected in a pure genetic drift model, as indicated by Waples (1989) test †.

LOCUS	Period		
	All 14 cycles of selection	7 cycles of HS selection	7 cycles of S_2 selection
ASG008	**	NS	NS
ASG024	**	*	*
ASG045	***	NS	***
ASG062	**	**	*
BNL05.10	NS	NS	**
BNL05.14A	***	***	NS
BNL05.47	NS	NS	*
BNL06.32	**	NS	NS
BNL07.26	*	NS	NS
BNL07.49	NS	NS	***
BNL07.71	**	*	*
BNL08.33	NS	NS	**
BNL09.11	NS	NS	***
BNL09.44	***	**	NS
BNL12.09	***	NS	NS
BNL13.05A	**	**	NS
BNL13.05B	*	*	NS
BNL15.07	*	NS	NS
CSU147	*	NS	NS
PHI10016	NS	NS	***
UMC005	**	**	NS
UMC007	*	NS	NS
UMC021	**	NS	*
UMC027	NS	NS	***
UMC030	***	***	*
UMC034	NS	NS	***
UMC036	NS	NS	***
UMC054	NS	*	NS
UMC055A	*	NS	NS
UMC060	***	NS	***
UMC062	*	**	NS

† NS means changes in allele frequencies were not significant by Waples (1989) test.

*,**,*** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively

Table 4. (continued)

LOCUS	Period		
	All 14 cycles of selection	7 cycles of HS selection	7 cycles of S_2 selection
UMC076	*	NS	NS
UMC080	NS	NS	**
UMC081	*	NS	*
UMC089	**	NS	*
UMC108	**	*	NS
UMC110	***	NS	NS
UMC113	*	NS	NS
UMC121	NS	NS	***
UMC128	**	*	NS
UMC131	NS	NS	NS
UMC137	*	NS	***
UMC147	*	NS	*
UMC158A	NS	NS	***
UMC158B	*	*	NS

Table 5. Means, standard errors (SE), and 95% confidence intervals for Nei (1978) genetic distances between cycles of selection of BSSS population. †

	N‡	Mean	SE	95% interval
BSSSP x BS13(S)C0	90	0.262	0.0045	(0.256 , 0.272)
- Non-RYD x BS13(S)C0	90	0.297	0.0047	(0.287 , 0.306)
- RYD x BS13(S)C0	90	0.303	0.0048	(0.293 , 0.312)
BSSSP x BS13(S)C7	102	0.328	0.0036	(0.321 , 0.335)
- Non-RYD x BS13(S)C7	102	0.358	0.0039	(0.350 , 0.366)
- RYD x BS13(S)C7	102	0.371	0.0038	(0.363 , 0.378)
BS13(S)C0 x BS13(S)C7	192	0.082	0.0025	(0.076 , 0.087)

† Including two group of lines, non-RYD and RYD, from BSSSP cycle.

‡ N stands for total number of estimated genetic distances.

Table 6. Means and standard errors (SE) for Nei (1978) genetic distances between progenitor lines and BSSSP, BS13(S)C0, and BS13(S)C7 populations.

Line	Nei's genetic distance.					
	BSSSP		BS13(S)C0		BS13(S)C7	
	Mean	SE	Mean	SE	Mean	SE
Non-RYD group						
A36-3-1-3	1.20	0.031	1.03	0.008	1.11	0.007
CI540	1.01	0.025	0.73	0.007	0.82	0.007
Ill. Hy	1.01	0.030	0.65	0.005	0.61	0.005
Ill.12E	0.92	0.019	0.72	0.007	0.77	0.005
LE23	0.94	0.031	0.65	0.007	0.70	0.006
Oh3167B	0.95	0.021	0.69	0.008	0.85	0.006
RYD group						
CI187-2	0.88	0.033	0.72	0.008	0.75	0.006
I159	0.96	0.039	0.75	0.008	0.85	0.006
I224	0.90	0.030	0.64	0.007	0.68	0.006
Ind.461-3	0.82	0.063	0.64	0.006	0.72	0.006
Ind.AH83	0.79	0.067	0.60	0.006	0.69	0.006
Ind.B2	0.85	0.042	0.57	0.006	0.69	0.005
Ind.Fe21073	0.94	0.042	0.81	0.008	0.88	0.005
Ind.Tr9-1-1-6	0.86	0.031	0.66	0.007	0.70	0.005
Os420	0.89	0.029	0.65	0.007	0.70	0.005
WD546	0.90	0.030	0.72	0.008	0.78	0.006
Mean all lines	0.93	0.015	0.70	0.003	0.77	0.003

Table 7. General com
means and standard e
that share a common

F1 hybrids group
Non-RYD
A36-3-1-3
CI540
Ill. Hy
Ill.12E
LE 23
Oh3167B
RYD
CI187-2
I159
I224
Ind.461-3
Ind.AH83

Table 7. General combining ability (GCA) yield effects for BSSSP progenitor lines, and means and standard errors (SE) for Nei (1978) genetic distances between F1 hybrids groups that share a common progenitor line and BSSSP, BS13(S)C0, and BS13(S)C7 populations.

F1 hybrids group	GCA for yield†	Average genetic distance					
	(q ha^{-1})	BSSSP		BS13(S)C0		BS13(S)C7	
		Mean	SE	Mean	SE	Mean	SE
Non-RYD							
A36-3-1-3	-0.63	0.18	0.012	0.52	0.012	0.60	0.014
CI540	6.14	0.16	0.009	0.45	0.013	0.52	0.014
Ill. Hy	4.77	0.16	0.007	0.40	0.008	0.42	0.009
Ill.12E	4.48	0.14	0.010	0.45	0.013	0.51	0.015
LE 23	-6.13	0.15	0.010	0.42	0.013	0.48	0.014
Oh3167B	-0.61	0.14	0.009	0.43	0.012	0.54	0.015
RYD							
CI187-2	-1.02	0.13	0.006	0.46	0.009	0.51	0.012
I159	3.18	0.14	0.006	0.42	0.010	0.47	0.012
I224	1.86	0.11	0.010	0.44	0.015	0.51	0.017
Ind.461-3	5.76	0.10	0.010	0.43	0.014	0.50	0.016
Ind.AH83	4.23	0.13	0.006	0.40	0.009	0.50	0.010
Ind.B2	-6.66	0.14	0.005	0.48	0.010	0.55	0.013
Ind.Fe21073	-5.99	0.13	0.008	0.44	0.012	0.49	0.013
Ind.Tr9-1-1-6	0.10	0.13	0.008	0.42	0.009	0.48	0.011
Os420	-3.57	0.14	0.007	0.46	0.009	0.52	0.012
WD546	-5.92	0.14	0.007	0.46	0.009	0.54	0.013
Mean all 120 F1 hybrids		0.14	0.003	0.44	0.005	0.51	0.006

† Estimates of GCA were obtained by Stucker and Hallauer (1992).

Table 8. Summary of genetic distances between F1 hybrids and BS13(S)C0 and BS13(S)C7 populations.

		BS13(S)C0	BS13(S)C7
F1 hybrids with lowest distances to BS13(S)CO			
Ill. Hy	LE23	0.35	0.36
LE23	I159	0.36	0.41
Ind.AH83	Ind.Fe21073	0.36	0.44
LE23	Ind.AH83	0.36	0.45
Ill. Hy	Ind.461-3	0.37	0.39
Mean		0.36	0.41
F1 hybrids with lowest distances to BS13(S)C7			
Ill. Hy	LE23	0.35	0.36
Ill. Hy	I159	0.39	0.39
Ill. Hy	Ind.Fe21073	0.39	0.39
Ill. Hy	Ind.461-3	0.37	0.39
Ill. Hy	Ind.Tr9-1-1-6	0.39	0.39
Mean		0.38	0.38
Mean all 120 F1 hybrids		0.42	0.49

Table 9. Correlation for genetic distances between lines and BSSSP population estimated on individuals of BS13(S)C0 and BS13(S)C7 populations.

Line	Correlation for genetic distances between lines and BSSSP	
	BS13(S)C0	BS13(S)C7
A36-3-1-3	0.51***	0.66***
CI540	0.62***	0.52***
Ill. Hy	0.36***	0.62***
Ill.12E	0.61***	0.53***
LE23	0.68***	0.65***
Oh3167B	0.70***	0.68***
CI187-2	0.69***	0.55***
I159	0.66***	0.60***
I224	0.57***	0.57***
Ind.461-3	0.61***	0.78***
Ind.AH83	0.70***	0.78***
Ind.B2	0.68***	0.59***
Ind.Fe21073	0.60***	0.61***
Ind.Tr9-1-1-6	0.77***	0.64***
Os420	0.70***	0.68***
WD546	0.71***	0.71***

*** Significant at 0.001 probability level.

Table 10. Correlation between genetic distances estimated in different cycles of selection, between genetic distances and general combining ability for yield, and between genetic distances and F1 grain yield. †

Variable 1	Variable 2	N‡	Correlation
NGD progenitor lines vs. BSSSP	NGD progenitor lines vs. BS13(S)C0	16	0.83***
NGD progenitor lines vs. BSSSP	NGD progenitor lines vs. BS13(S)C7	16	0.72**
NGD progenitor lines vs. BS13(S)C0	NGD progenitor lines vs. BS13(S)C7	16	0.93***
NGD F1 hybrids groups vs. BSSSP	NGD F1 hybrids groups vs. BS13(S)C0	16	0.50*
NGD F1 hybrids groups vs. BSSSP	NGD F1 hybrids groups vs. BS13(S)C7	16	0.32
NGD F1 hybrids groups vs. BS13(S)C0	NGD F1 hybrids groups vs. BS13(S)C7	16	0.86***
NGD F1 hybrids groups vs. BSSSP	General combining ability for yield	16	-0.12
NGD F1 hybrids groups vs. BS13(S)C0	General combining ability for yield	16	-0.37
NGD F1 hybrids groups vs. BS13(S)C7	General combining ability for yield	16	-0.30
NGD F1 hybrids vs. BS13(S)C0	F1 grain yield	96	-0.17
NGD F1 hybrids vs. BS13(S)C7	F1 grain yield	96	-0.16

† Estimates for GCA and F1 grain yield were obtained by Stucker and Hallauer (1992).

‡ Sample size.

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

Table 11. Nei's genetic distance among 16 lines of BSSSP population.

Line	Code	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13	L14	L15	L16
A3G-3-1-3	L1		1.20	1.32	1.04	1.04	1.06	1.16	1.13	1.25	1.36	1.34	1.37	1.10	1.15	1.17	1.36
CI540	L2			1.12	0.84	1.08	1.13	1.02	1.02	0.97	1.00	0.93	1.00	0.95	0.90	0.92	1.03
Ill. Hy	L3				1.02	1.14	0.99	1.01	0.93	0.94	0.91	0.82	1.05	0.97	0.93	1.00	1.06
Ill.12E	L4					0.86	0.90	0.92	0.97	0.93	0.89	0.81	0.93	0.78	0.99	0.93	0.91
LE23	L5						0.94	0.84	1.01	0.96	0.91	0.88	0.97	0.68	0.81	0.97	1.04
Oh3167B	L6							0.87	1.03	0.85	0.85	0.89	0.96	0.94	0.98	0.94	0.84
CI187-2	L7								0.78	0.81	0.64	0.78	0.90	0.82	0.77	1.00	0.93
I159	L8									0.78	0.72	0.76	0.89	0.93	0.87	0.82	0.92
I224	L9										0.11	0.78	0.64	0.76	0.87	0.78	0.94
Ind.461-3	L10											0.63	0.71	0.82	0.80	0.74	0.82
Ind.AH83	L11												0.97	0.88	0.78	0.78	0.69
Ind.B2	L12													0.80	1.00	0.97	0.97
Ind.Fe21073	L13														0.72	0.78	1.03
Ind.Tr9-1-1-6	L14															0.84	0.96
Os420	L15																0.83
WD546	L16																

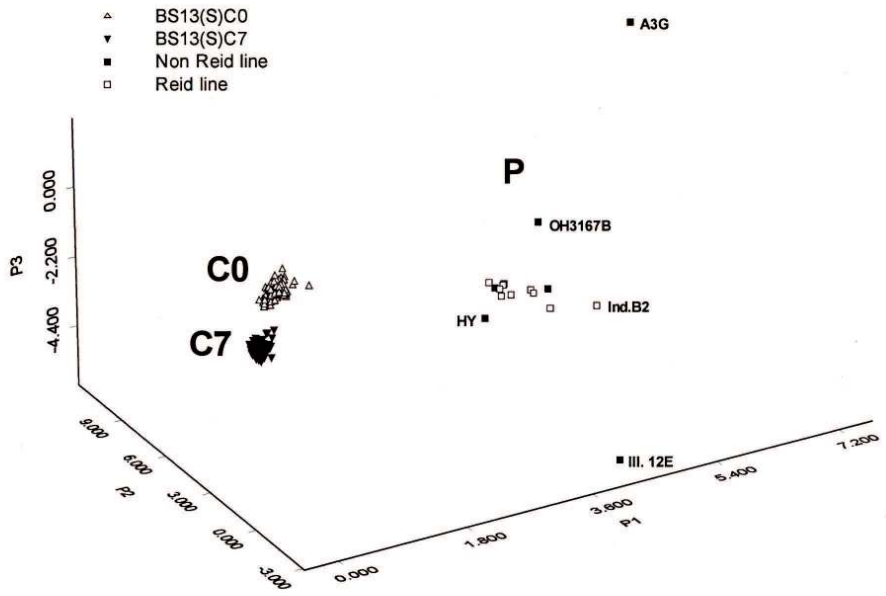

Figure 1. Diagram from the first 3 principal components (P1, P2, and P3) scores of individuals from BSSSP, BS13(S)C0, and BS13(S)C7 populations at 105 RFLP loci. Progenitor lines from BSSSP traced to Yellow Reid Dent and non Yellow Reid Dent origins are marked with different symbols. Progenitor lines with highest distance from other lines are identified in the diagram.

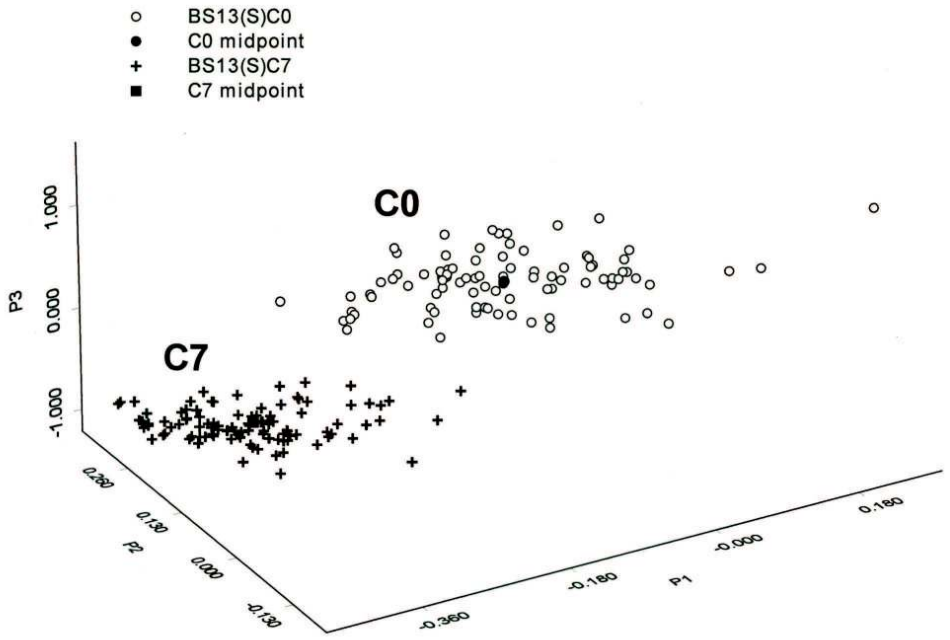

Figure 2. Diagram from the first 3 principal components (P1, P2, and P3) scores of individuals from BS13(S)C0 and BS13(S)C7 populations.

GENERAL CONCLUSIONS

General Discussion

In this study, 105 RFLP loci were used to determine the effects of 14 cycles of recurrent selection on genetic structure of BSSS population. Results indicated that the original synthetic, BSSSP, had a significant amount of genetic variation. BSSP showed high mean expected heterozygosity, intermediate average for the most common allele, low proportion of monomorphic loci, half of the alleles traced to one or two progenitor lines, high mean number of alleles per locus, and high dispersion of progenitor lines based on PCA analysis. These findings are also consistent with previous studies that reported significant genetic variation for this synthetic based on phenotypic data (Stucker and Hallauer, 1992) and RFLP data (Neuhausen, 1989; Messmer et al., 1991; Labate et al., 1997). This high genetic variation, observed herein and in other studies, is also supported by the investigations of Eberhart et al. (1973), Lamkey (1992), and Holthaus and Lamkey (1995) that reported substantial genetic gains for grain yield and other traits during the first seven cycles of recurrent in this population based on HS-progeny selection.

High proportion of rare alleles was an important component on the genetic variation in BSSSP and most of the alleles lost over the cycles of selection were the ones that had low frequency in the BSSP population. This finding is in agreement with theoretical expectations (Kimura, 1957) and is useful to illustrate the point that rare neutral and effectively neutral alleles tend to be lost, mainly under low effective size population. A large proportion of genes with low selective values is more probable for traits that are controlled by a high number of genes (Falconer and Mackay, 1996; Hartl and Clark, 1997). Since the recurrent program in BSSSP has been characterized by average low effective population size, many

rare BSSP alleles, with weak coefficient of selection for the quantitative traits being selected, that are assumed to be controlled by a high number of genes, were or are going to be lost over the cycles of selection. Further studies could check this inference by tracking the frequencies of alleles linked to QTLs for the traits of interest over the cycles of recurrent selection in this population.

Genetic distances among parental lines were not good predictors of grain yield single-crosses performance. A low and nonsignificant correlation of genetic distance interlines (data obtained in this study) and F_1 hybrid performance (data obtained by Stucker and Hallauer, 1992) reported herein is in agreement with theoretical expectations (Bernardo, 1992) and the majority of empirical studies (Lamkey et al., 1987, Lee et al., 1989; Godshalk et al., 1990; Melchinger et al., 1990a and 1990b; Dudley et al., 1991; Ajmone Marsan et al., 1998). Thus suggests the need for further studies in this subject and indicates that predicting F_1 hybrid performance for quantitative traits based exclusively on genetic distances interlines of random sample of marker loci has low practical value for maize breeding programs.

There was a significant reduction of genetic variability over the cycles of selection. This conclusion is based on the observation that all the diverse measures of genetic variation (proportion of polymorphic loci, mean expected heterozygosity, mean number of alleles, average frequency of the most common allele, and dispersion of individuals within cycles of selection based on PCA analysis) detected great loss of variation over cycles of selection. This bottleneck effect observed herein follows what is expected to occur when there is a low effective population size (Nei, 1987; Hedrick, 1999), meaning that small number of progenies selected in each cycle, need to be considered as an important factor in reducing the genetic variability over cycles of selection.

Both, genetic distance and PCA results, gathering evidence that there was a considerable increase of genetic distance among cycles of selection. Low effective population size, high initial genetic variation, and high number of generations are factors that are directly proportional to the magnitude of loss of genetic variation within and increasing genetic variation among subpopulations over time (Hedrick, 1999). All these conditions were observed in this investigation for the RFLP loci we sampled and can explain the increase in the genetic divergence among the cycles of selection and also the loss of genetic variation within cycles of selection in this long-term maize recurrent selection program.

The greatest reduction in genetic variability occurred in the HS selection period. All measures of genetic variation within and among cycles of selection indicated the genetic variability in BS13(S)C0 was more similar to BS13(S)C7, rather than being intermediate between BSSSP and BS13(S)C7. These findings can be explained considering that the HS selection was characterized by a lower effective population size and had a founder population with much more variability than the S_2-selection period.

The next six cycles of selection (C9, C11, and C13) are predicted to show low mean expected heterozygosity values; values that are predicted to be similar to the estimated values for the BS13(S)C7 population. These results and the lower response for grain yield in the first six cycles of S_2-selection (Holthaus and Lamkey, 1995) suggest that BS13(S)C7 can be a founder population with inadequate genetic variability to sustain substantial long-term genetic gains per se for grain yield. Further evaluations could check this inference by evaluating the next cycles of selection in this population, based on phenotypic and molecular data.

In general, the genotypic frequencies in BS13(S)C0 and BS13(S)C7 were consistent with Hardy-Weinberg ratios, indicating observed and expected heterozygosity estimates were in close agreement in each population for most of the loci. As expected, all loci in BSSSP were not in H-W equilibrium. The inbreeding coefficient, $F_{IS,}$ was similar to one for BSSSP and nonsignificant and close to zero for BS13(S)C0 and BS13(S)C7. These findings are consistent with expectations, because the samples of BS13(S)C0 and BS13(S)C7 were obtained after two generations of random mating and BSSSP is a group of highly inbred lines.

Excess of homozygosity was the main reason for deviations from Hardy- Weinberg proportions in 10 loci in BS13(S)C0 and seven in BS13(S)C7. Positive assortative mating among individuals with similar flowering time and/or contamination with pollen or seed may explain the excess of homozygotes observed herein in some loci. Positive assortative mating was also indicated in other maize studies as a probable cause of observed excess of homozygosity (Dubreuil and Charcosset, 1998; Labate et al., 2000).

The results of this study suggested that selection, directly or by hitchhiking, was also an important factor in changing allele frequencies over the cycles of selection for the RFLP loci we sampled. Considering the 14 cycles of selection period, about 30% of the loci showed changes in allele frequencies higher than could be expected only by genetic drift. Ten of these loci also showed similar patterns of changing in allele frequencies in the reciprocal recurrent selection involving BSSS and BSCB1 populations (Labate et al., 1999). Could be interesting to evaluate the effects of these loci on the response for the traits undergoing selection in a marker assisted selection program based on the original BSSS population. Also, future research could address the question of which loci were linked with QTLs over

cycles of selection for the traits of interest via evaluation of progenies from different cycles of selection based on molecular and phenotypic data.

The divergence of genotypes (lines, F1 hybrids, and groups of F1 hybrids) was positively associated with their initial genetic divergence to BSSP population. Results obtained in this study suggested that line Ill. Hy may have exhibited a selective advantage over the cycles of selection. It showed one of the highest genetic distances to BSSSP, but the lowest genetic distance to BS13(S)C7 and a lower than average genetic distance to BS13(S)C0. Moreover, five of its unique alleles had their frequencies increased higher than could be expected by genetic drift alone.

The results obtained herein and also in other investigations indicated that molecular markers are useful tools to evaluate maize recurrent selection programs. However, much more knowledge is generated when molecular and phenotypic variation are evaluated jointly over the cycles of selection. If different programs of recurrent selection were evaluated based on phenotypic data and also on a standard large set of marker loci, great advances could be obtained in the study of populations, breeding methods, traits undergoing selection, environments, and marker loci. Such knowledge could lead to the design of more efficient maize recurrent selection programs based on phenotypic and marker assisted selection.

References

Ajmone Marsan, P., P. Castiglioni, F. Fusari, M. Kuiper, and M. Motto. 1998. Genetic diversity and its relationship to hybrid performance in maize as revealed by RFLP and AFLP markers. Theor. Appl. Genet. 96: 219-227.

Bernardo, R. 1992. Relationship between single-cross performance and molecular marker heterozygosity. Theor. Appl. Genet. 83: 628-634.

Dubreil, P., and A. Charcosset. 1998. Genetic diversity within and among maize populations: a comparison between isozyme and nuclear RFLP loci. Theor. Appl. Genet. 96: 577-587.

Dudley, J.M., M.A. Saghai Maroof, and G.K. Rufener. 1991. Molecular markers and grouping of parents in a maize breeding program. Crop Sci. 31: 718-723.

Eberhart, S.A., S. Debela, and A.R. Hallauer. 1973. Reciprocal recurrent selection in the BSSS and BSCB1 maize populations and half-sib selection in BSSS. Crop Sci. 13: 451-456.

Falconer, D.S., and T.F.C. Mackay. 1996. Introduction to quantitative genetics. 4th ed. Longman, England.

Godshalk, E.B., M. Lee, and K.R. Lamkey. 1990. Relationship of restriction fragment length polymorphisms to single-cross hybrid performance of maize. Theor. Appl Genet. 80:273-280.

Hartl, D.L., and A.G. Clark. 1997. Principles of population genetics. 3rd ed. Sinauer, Massachusetts.

Hedrick, P.W. 1999. Perspective: highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313-318.

Holthaus, J.H., and K.R. Lamkey. 1995. Populations means and genetic variances in selected and unselected Iowa Stiff Stalk Synthetic maize populations. Crop Sci. 35: 1581-1589.

Kimura, M. 1957. Some problems of stochastic processes in genetics. Ann. Math. Stat. 28: 882-901.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1997. Molecular genetic diversity after reciprocal recurrent selection in BSSS and BSCB1 maize populations. Crop Sci. 37:416-423.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 1999. Temporal changes in allele frequencies in two reciprocally selected maize populations. Theor. Appl. Genet. 99: 1166-1178.

Labate, J.A., K.R. Lamkey, M. Lee, and W.L. Woodman. 2000. Hardy- Weinberg and linkage equilibrium estimates in the BSSS and BSCB1 random mated populations. Maydica 45: 243-255.

Lamkey, K.R. 1992. Fifty years of recurrent selection in the Iowa Stiff Stalk Synthetic maize populations. Maydica 37: 19-28

Lamkey, K.R, A.R. Hallauer, and A.L. Kahler. 1987. Allelic differences at enzyme loci and hybrid performance in maize. Journal of Heredity 78: 231-234.

Lee, M., E.B. Godshalk, K.R. Lamkey, and W.W. Woodman. 1989. Association of restriction fragment length polymorphisms among maize inbreds with agronomic performance of their crosses. Crop Sci. 29: 1067-1071.

Melchinger, A.E., M. Lee, , K.R. Lamkey, and W.L. Woodman. 1990a. Genetic diversity for restriction fragment length polymorphisms: relation to estimated genetic effects in maize hybrid. Crop Sci. 30: 1033-1040

Melchinger, A.E., M. Lee, K.R. Lamkey, A.R. Hallauer, and W.L. Woodman. 1990b. Genetic diversity for restriction fragment length polymorphisms and heterosis for two diallel sets of maize inbreds. Theor. Appl. Genet. 80: 488-496.

Messmer, M.M., A.E. Melchinger, M. Lee, W.L. Woodman, E.A. Lee, and K.R. Lamkey. 1991. Genetic diversity among progenitors and elite lines from the Iowa Stiff Stalk Synthetic (BSSS) maize populations: comparison of allozyme and RFLP data. Theor. Appl. Genet. 83: 97-107.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

Neuhausen, S.L. 1989. A survey of Iowa Stiff Stalk Synthetic parents, derived inbreds, and BSSS(HT)C5 using RFLP analysis. Maize Gen. Coop. Newsl. 63: 110-111.

Stucker, D.S. and A.R. Hallauer. 1992. Genetic variability as affected by selection in Iowa Stiff Stalk Synthetic maize. Journal of Heredity 83: 410-418.