

J.W. DUDLEY*

From Means to QTL: The Illinois Long- Term Selection Experiment as a Case Study in Quantitative Genetics

Dep. of Crop Sciences, Univ. of Illinois, Urbana, IL 61801.
This research was supported by the Illinois Agric. Exp. Stn. and a
grant from Renessen, LLC. Received 9 July 2007.

*Corresponding author (jdudley@uiuc.edu).

Published in Crop Sci. 47(S3) S20–S31 (2007).
doi: 10.2135/cropsci2007.04.0003IPBS
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA





Divergent selection for oil and protein concentration in the corn (*Zea mays* L.) kernel was initiated at the University of Illinois in 1896 by C.G. Hopkins. In 2005, 106 generations of selection had been completed for high oil and 105 for high protein. Limits to selection for low oil and low protein were reached but not for high oil or high protein. Over the more than 100 yr of the existence of the program a number of attempts have been made to analyze the experiment using quantitative genetic tools. The purpose of this paper is to trace the use of quantitative genetic techniques to analyze the results of divergent long-term selection for oil and protein, to relate results to the question of the need for divergent parents for quantitative trait locus (QTL) analysis, and to provide a new look at the reasons for long-continued progress from selection. Key findings include (i) progress from selection was much greater than could have been predicted; (ii) based on both classical quantitative genetic analysis and QTL studies, a large number of QTL are involved in control of the three traits; (iii) the number of QTL identified in a given study cannot be predicted by the magnitude of genetic variance or the divergence of the parents but is a function of the number of markers used and the number of lines evaluated; and (iv) epistasis may be an important factor in explaining long-term response to selection.

Divergent long-term selection is a tool for understanding genetic control of a trait and for concentrating alleles for contrasting versions of the trait. The Illinois long-term selection experiment for kernel oil and protein concentration is a classic example of use of divergent selection to understand genetic control of traits. The history of this experiment was reviewed extensively in the proceedings of a symposium on long-term selection (Dudley and Lambert, 2004). Details of selection intensity, breeding procedure, and methods of chemical analysis are found in those proceedings. Briefly, the experiment was started by C.G. Hopkins in 1896 (Hopkins, 1899). The open-pollinated corn cultivar Burr's White was the founder population. Four selected strains were established: Illinois High Oil (IHO), Illinois Low Oil (ILO), Illinois High Protein (IHP), and Illinois Low Protein (ILP). Selection in each strain was in the direction indicated by the name of the strain (e.g., IHO was selected for high oil concentration in the kernel). Mass selection, based on analysis of individual ears, was used with a selection intensity of approximately one ear out of five (12 ears out of 60 most recently) for most of the experiment (Dudley and Lambert, 2004). After 48 generations, reverse selection was started in each strain to establish the Reverse High Oil (RHO), Reverse Low Oil (RLO), Reverse High Protein (RHP), and Reverse Low Protein (RLP) strains (Woodworth et al., 1952). After seven generations of selection in RHO, selection was again reversed to create the Switchback High Oil strain (SHO). The objective of the reverse selection strains was to determine whether genetic variability had been exhausted in the selected strains.

The tools of quantitative genetics include means, regression analysis, variance estimation (particularly genetic variances), genetic correlations, mating designs, random mating, and more recently QTL analysis. The purpose of this paper is to trace the use of quantitative genetic techniques to analyze the results of divergent long-term selection for oil and protein concentration in the corn (*Zea mays* L.) kernel at the University of Illinois, to relate results to the question of the need to use divergent parents for QTL analysis, and to provide a new look at the reasons for long-continued progress from selection.

Results

The results are discussed under the major headings of Classical Quantitative Genetics, QTL Studies, Explanations of Long-Term Response, and Conclusions. The materials and methods used in the studies discussed are in the original papers.

Abbreviations: IHO, Illinois High Oil; IHP, Illinois High Protein; ILO, Illinois Low Oil; ILP, Illinois Low Protein; QTL, quantitative trait locus (loci); RHO, Reverse High Oil; RHP, Reverse High Protein; RLO, Reverse Low Oil; RLP, Reverse Low Protein; SHO, Switchback High Oil; SIM, simple interval mapping; SMA, single marker analysis; SNP, single nucleotide polymorphism.

Classical Quantitative Genetics

Means

For the duration of the experiment, means of ears sampled each cycle are plotted against generations in Fig. 1 and 2 for the oil and protein strains, respectively. The earliest report of results from this experiment (Hopkins, 1899) compared means of IHO cycle 2 with ILO cycle 2 for oil and IHP cycle 2 with ILP cycle 2 for protein. For the IHO-ILO comparison, a difference of 11.5 g kg⁻¹ oil was found while for the IHP-ILP comparison, the difference in protein was 5.0 g kg⁻¹. From these results Hopkins concluded that selection for oil and protein was effective even though there was no test of significance of the observed differences. After 10 generations of selection Smith (1908) found a difference of 56.2 g kg⁻¹ protein between IHP and ILP and a difference of 47.1 g kg⁻¹ oil between IHO and ILO. These results were based on means of the ears grown in 1906 in each strain and again there was no statistical test of the significance of differences between means. After 28 generations of selection the difference in means between IHO and ILO was 83.5 g kg⁻¹ oil and the difference between IHP and ILP was 82.2 g kg⁻¹ protein (Winter, 1929). After 50 generations of selection the difference between IHO and ILO was 143.5 g kg⁻¹ oil and the difference between IHP and ILP was 145.4 g kg⁻¹ protein (Woodworth et al., 1952). Results of two cycles of reverse selection (selection within a strain in the opposite direction, e.g., selection for low protein in IHP) showed apparent progress in RHO, RHP, and RLP but little progress in RLO.

Leng (1962) attempted to apply the results of 61 generations of selection to breeding theory. At 61 generations, the difference between IHO and ILO was 140.6 g kg⁻¹ oil (34 times the phenotypic standard error of the original population) and the difference between IHP and ILP was 197.5 g kg⁻¹ protein (19 times the phenotypic standard error in the original population). The mean values of the strains were beyond values found in any nonselected strains of corn. Thus Leng concluded that breeding systems which permit continued recombination during many generations may be especially effective in obtaining maximum levels of expression of desired traits.

The first comparison of different cycles of the selection experiment grown in the same environments was reported by Dudley et al. (1974). Generations 65 to 70 of the forward selection strains were compared as were generations 19 to 22 of the reverse selection strains. Earlier generations were not included because seed was not available. Progress was significant and in the appropriate direction for all strains.

At generation 90, progress measured in replicated trials comparing the last 20 to 25 generations of selection of the forward and reverse strains was similar to that measured in each individual cycle (Dudley and Lambert 1992). No progress in either ILO or ILP was observed after generation 76. After 100 generations, results of replicated trials compar-

ing the last 33 generations of selection showed continued progress for oil in IHO, RHO, RLO, and SHO (Dudley and Lambert 2004) and for protein in IHP, RHP, and RLP.

Means are an important measure of progress and an important statistical tool in all the work which has been done to help understand the genetic control of oil and protein. Mean values for individual ears were the units of selection. Means have been used in conjunction with other quantitative genetic tools in work following that of Smith (1908). Thus, further discussion of means will be included with discussion of variances or other quantitative genetic tools.

Variances

Winter (1929) was the first to report the effect of divergent selection on standard errors and coefficients of variation (CVs). He showed that selection in the high direction for either oil or protein decreased the CV among ears within strains while selection in the low direction increased it. Conversely, selection in the high direction increased the among ears within strains standard deviation while selection in the low direction decreased it. Based on these results, Winter (1929) concluded that continued progress in the high direction should be possible for both oil and protein but that a selection limit was being approached in the low strains. Woodworth et al. (1952) noted the same trends in standard errors and CVs within strains as reported by Winter (1929).

The first estimates of genetic variance in the selected strains were from a half-sib mating design in generation 65 of IHO, ILO, IHP, and ILP (Dudley and Lambert, 1969). Significant genetic variance was found for both oil and protein in all four strains. Genetic variance in IHO was larger than in ILO for both oil and protein with the difference being larger for oil. Similarly the genetic variance in IHP was larger for both oil and protein than in ILP. Predicted gain for one cycle of 20% selection for the selected trait in all four strains was similar to the average gain per generation for the preceding 65 generations.

Genetic variances among S_1 lines were estimated for the earliest (generation 69) and latest (generation 98) cycles available of IHP, IHO, and SHO (Dudley and Lambert, 2004). No significant differences in genetic variances between generations 69 and 98 were found for the selected trait in any of the

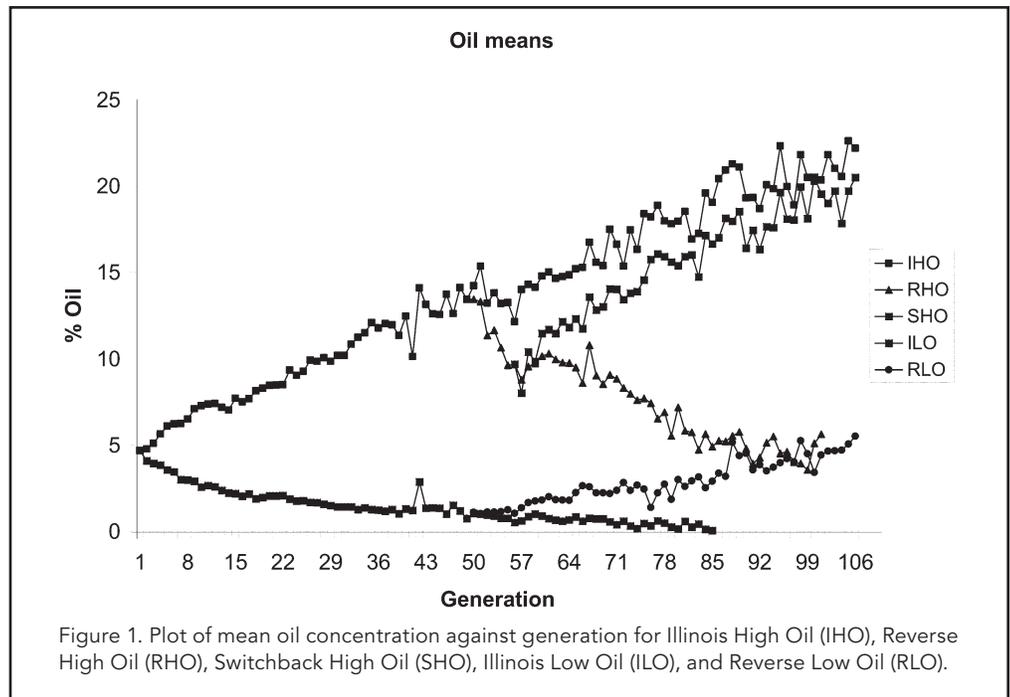


Figure 1. Plot of mean oil concentration against generation for Illinois High Oil (IHO), Reverse High Oil (RHO), Switchback High Oil (SHO), Illinois Low Oil (ILO), and Reverse Low Oil (RLO).

strains tested (Dudley and Lambert, 2004) although significant genetic variance was found in all three strains.

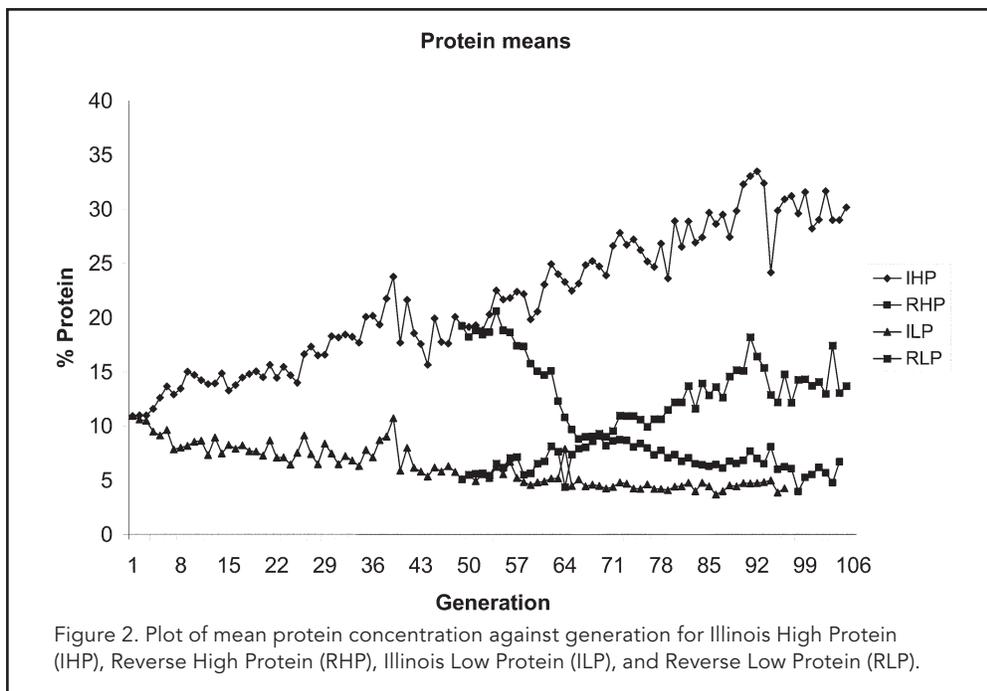
Random Mating

Moreno-Gonzalez et al. (1975) reported estimates of genetic variance obtained from a Design III study of the F_2 and random-mated F_6 generations of the cross of generations 70 of IHO \times ILO. In accordance with quantitative genetic theory, estimates of additive genetic variance in the F_6 were significantly lower than in the F_2 indicating the build-up of coupling phase linkages for alleles for high oil in IHO and for alleles for low oil in ILO. Estimates of dominance variance were similar in the F_2 and F_6 suggesting a lack of effect of linkage on dominance variance.

In a similar Design III study of the F_2 and F_6 generations of the cross of generations 70 of IHP \times ILP (Dudley 1994), additive genetic variance for protein was significantly less in the F_6 than in the F_2 suggesting the build-up of coupling phase linkages for high protein alleles in IHP and for low protein alleles in ILP. The measurement of marker-QTL associations in the F_2 and in subsequent random-mated generations was proposed as a method of identifying tight marker-QTL associations.

Regression

Dudley et al. (1974) first reported use of regression of means on cycles of selection to measure rate of progress through generation 70. The experiment was divided into segments based on changes in breeding procedure and the selection environment. Response was measured by regression of means on cycles within segments. Realized heritabilities were calculated for each segment as regression of means on cumulative selection differentials. Results of this process



also were reported after 90 (Dudley and Lambert, 1992) and 100 (Dudley and Lambert, 2004) generations. The change per generation for oil in IHO per generation ranged from $2.2 \pm 0.7 \text{ g kg}^{-1}$ oil in the first nine generations to $1.6 \pm 0.1 \text{ g kg}^{-1}$ in generations 59 to 100 (Dudley and Lambert, 2004). The change per generation in generations 0 to 9 of ILO ($-2.1 \pm 0.3 \text{ g kg}^{-1}$ oil) was very similar to IHO but from generations 59 to 88 the change was a nonsignificant $-0.1 \pm 0.4 \text{ g kg}^{-1}$ oil per generation. In contrast, the realized heritability in generations 0 to 9 was 0.50 for both IHO and ILO and in generations 59 to 88 realized heritability was 0.08 for both strains. Inbreeding coefficients in generation 70, calculated based on effective population sizes each generation, were greater than 0.8 for all strains.

Number of Effective Factors

Dudley (1977) evaluated theoretical limits to selection based on additive genetic variance, frequency of alleles favorable for high oil or high protein (q), and number of effective factors. From this study, estimates of 54 effective factors differentiating IHO and ILO for oil and 122 effective factors differentiating IHP and ILP for protein in generation 76 were obtained. Estimates of q were 0.25 for oil and 0.37 for protein in the original Burr's White. Based on these estimates, the observed progress from selection was well within what might be expected with relatively low frequencies of favorable alleles in the original population and a reasonable number of segregating loci. At generation 90, gene frequency estimates for the original population (0.22 for oil and 0.32 for protein) (Dudley and Lambert, 1992) were slightly lower than estimated by Dudley (1977). Because additional progress had been made in IHO and IHP, estimates of effective factors increased to 69 for oil and 173 for protein. As expected, with continued progress

in the high direction, estimates of q for oil (0.20) and for protein (0.24) in the original Burr's White were slightly lower after 100 generations (Dudley and Lambert, 2004) than when estimated after 90 generations.

Correlated Response

Each generation a sample of the ears from the strains selected for oil was analyzed for protein and a sample of the strains selected for protein was analyzed for oil (Dudley et al., 1974). From results of the first 70 generations, protein concentration increased in both IHO and ILO from 109 to approximately 150 g kg^{-1} . In the IHP strain, oil

did not change whereas in the ILP strain oil concentration decreased from 47 g kg^{-1} in generation 0 to 31 g kg^{-1} in generation 70. Replicated trials of selected generations from generation 65 to 100 in IHO showed a significant increase from 148 to 158 g kg^{-1} protein and a decrease in starch from 435 to 336 g kg^{-1} as oil increased from 161 to 224 g kg^{-1} (Dudley and Lambert, 2004). Similar results were obtained in the latest generations of RHO and SHO. In IHP, there was no significant change in oil from generation 67 to 99 but starch decreased from 327 to 234 g kg^{-1} as protein increased from 228 to 269 g kg^{-1} . Similar results were obtained in RHP and RLP in that oil did not change and starch decreased as protein increased and increased as protein decreased.

Estimates of phenotypic correlations were available for a number of studies of crosses of IHO \times ILO and IHP \times ILP. In general, correlations of starch with protein were highly negative (ranging from -0.82 to -0.96) in the IHP \times ILP cross regardless of random-mated generation (Table 1). Correlations of starch with protein in the IHO \times ILO cross were lower (varying from -0.21 to -0.73). The correlations of starch with oil were similar in both crosses with a tendency for the RM7 and RM10 generations to have correlations of greater magnitude than for the earlier random-mated generations. The correlations of protein with oil were generally low but positive in all random-mated generations of both crosses. Genetic correlations were similar to phenotypic correlations. These results agree with the results from the observed means in predicting that selection for higher protein will result in decreased starch with little change in oil and selection for higher oil tends to reduce starch and may result in a small increase in protein.

Classical Quantitative Genetics Summary

Use of classical quantitative genetic methods demonstrated real progress from selection over 100 generations. Heritability estimates from replicated trials were high although realized heritability estimates varied in different phases of the experiment (Dudley and Lambert, 2004). Estimates of genetic variance in random-mated and non-random-mated generations demonstrated the build-up of coupling phase linkages for both oil and protein. Limits to selection were reached in both ILO and ILP but not in IHP or IHO. Genetic control of oil and protein was projected as being due a large number of genes segregating in the original population with the frequency of the favorable alleles being relatively low in the original Burr's White. These results lead to the conclusion that progress from selection was much larger than what might ordinarily be expected based on genetic variability in an original segregating generation. Information on correlated response from means of selected strains, comparisons of cycles in replicated trials, and correlations in different random-mated generations suggest it will be difficult to increase either protein or oil concentration without decreasing starch concentration and that oil and protein have a small positive correlation.

QTL Studies

The QTL studies aimed at identification of significant marker-QTL associations in progenies resulting from crosses of divergently selected strains have encompassed a wide range of types and numbers of progenies as well as types and numbers of markers. In most of the studies, classical quantitative genetic information such as means, genetic variances, and heritabilities was available and used to help interpret the data. Emphasis in several studies was on use of random mating to identify tight linkages between markers and QTL. All studies discussed used progenies from the cross of either IHO × ILO or IHP × ILP.

The first study reported in the literature involved 100 F₃ progenies from the cross of generation 76 of IHP × ILP and 90 restriction fragment length polymorphism markers (Goldman et al., 1993; Goldman et al., 1994). Twenty-two significant marker QTL associations were found for protein, 19 for starch, and 25 for oil. The number of significant marker-QTL associations for oil (Goldman et al., 1994) was similar to the number found for protein by Goldman et al. (1993) despite the fact that the parents (IHP and ILP) were more divergent for protein than

for oil and the genetic variance for protein would have been expected to be higher for protein than for oil.

Random Mating and QTL Identification

The first report comparing marker-QTL associations in non-random-mated and subsequent random-mated generations was by Dudley et al. (2004). Five to seven plants from IHP generation 70 were crossed to five to seven plants from ILP generation 70. The Syn0 (F₂) was random-mated four times to produce the Syn4. Two hundred S₁ lines each from the Syn0 and the Syn4 generations were evaluated as lines per se and as testcrosses to two inbred testers. Means for starch, protein, and oil of the S₁ and testcrosses were similar in the Syn0 and Syn4 for each trait. In accordance with theory and the results of Dudley (1994), genetic variances for protein and starch were significantly less in the Syn4 than in the Syn0 for all progeny types suggesting the build-up of coupling phase linkages for protein and starch.

Dudley et al. (2004) hypothesized that the proportion of significant marker-QTL associations should be higher in the Syn0 than in the Syn4 and that marker-associated effects should be larger in the Syn0 than in the Syn4. For protein, there were 38, 32, and 32 significant marker-associated effects (for the S₁, testcross to FR1064, and testcross to FR616 progenies respectively) identified by single marker analysis (SMA) in the Syn0 and 13, 7, and 3 in the Syn4. Similar results were obtained for starch. Because 64 markers were used, the proportions of significant markers were approximately 0.5 in the Syn0 and <0.2 in the Syn4. Thus random mating reduced the proportion of significant

Table 1. Phenotypic correlations among protein, oil, and starch for per se progenies from different experiments.

Experiment	Generation	Population	Starch vs. protein	Starch vs. oil	Protein vs. oil
IHO × ILO [†]					
Willmot et al., 2006	Syn0	Per se	-0.69**	-0.39**	0.30**
		Testcross	-0.43**	-0.42**	0.09
	Syn4	Per se	-0.62**	-0.21**	0.09
		Testcross	-0.21**	-0.33**	-0.11
Clark et al., 2006	RM10	Per se	-0.65**	-0.55**	0.23**
		Testcross	-0.73**	-0.75**	0.18**
IHP × ILP [‡]					
Dudley et al., 2004	Syn0	Per se	-0.95**	-0.25**	0.15*
		Testcross	-0.82**	-0.41**	0.19**
	Syn4	Per se	-0.94**	-0.25**	0.14
		Testcross	-0.96**	-0.50**	-0.22**
Dudley et al., 2007	RM7	Per se	-0.92**	-0.40**	0.05
		Testcross	-0.91**	-0.70**	0.41**

*Significant at the 0.05 probability level.

**Significant at the 0.01 probability level.

[†]IHO, Illinois High Oil; ILO, Illinois Low Oil.

[‡]IHP, Illinois High Protein; ILP, Illinois Low Protein.

Table 2. Comparison of IHP × ILP RM7 data with IHO × ILO RM10 data.[†]

Statistic	Population	Oil		Protein		Starch	
		IHO × ILO	IHP × ILP	IHO × ILO	IHP × ILP	IHO × ILO	IHP × ILP
Mean, g kg ⁻¹	Per se	70.6	44.1	129	134	585	681
	Testcross	49.2	45.3	126	103	677	699
Genetic variability	Per se	119.0**	9.88**	71.7**	219.0**	716**	114**
	Testcross	10.2**	3.03**	21.4**	32.0**	33.2**	23.2**
Heritability	Per se	0.93	0.89	0.88	0.92	0.84	0.94
	Testcross	0.94	0.77	0.77	0.83	0.83	0.79
Error [‡]	Per se	109.7	8.7	61.3	199.8	608.2	105.5
	Testcross	9.5	2.3	15.8	25.5	27.3	17.8
Avg. efficiency [§]	Per se	0.53	0.14	0.38	0.76	0.53	1.68
	Testcross	0.18	0.07	0.21	0.25	0.18	0.62
CV, %	Per se	9.7	5.2	6.1	6.7	4.4	1.0
	Testcross	4.5	4.5	5.7	5.7	1.0	0.8

**Significant at the 0.01 probability level.

[†]IHO, Illinois High Oil; ILO, Illinois Low Oil; IHP, Illinois High Protein; ILP, Illinois Low Protein.

[‡]Average (over all markers) of among families within genotypes error used to test significance of marker effect.

[§]Average absolute values of significant marker effects from simple interval mapping analysis.

marker-associated effects. In addition, for both the S₁ and testcross progenies, the mean absolute values of significant marker-associated effects for the Syn4 for protein and starch were approximately half those for significant effects in the Syn0. Unlike the results of Goldman et al. (1993, 1994) the number of significant marker-associated effects for oil was less than the number for protein or starch.

A study similar to that of the IHP × ILP cross was reported by Willmot et al. (2006) in the IHO × ILO cross. Five to seven plants from IHO generation 70 were crossed to five to seven plants from ILO generation 70. One hundred sixty-seven S₅ lines from the RM1 and 195 S₅ lines from the RM5 and their testcrosses to Illinois Foundation Seeds inbred FR1064 were evaluated. One hundred forty-seven simple sequence repeat markers were used. Unlike the study of Moreno-Gonzalez et al. (1975) there was a significant increase in mean oil concentration from the RM1 to the RM5 for both the lines per se and their testcross progenies. Heritability for oil was approximately 0.90 for both the RM1 and the RM5 for both lines per se and testcrosses. For oil, approximately two-thirds as many marker-associated effects were identified as significant by single marker analysis in the RM5 as were significant in the RM1 for both the lines per se and the testcrosses. Thus, in agreement with the results of Dudley et al. (2004) for protein in the cross of IHP × ILP, random mating reduced the number of significant marker-QTL associations.

Extensive QTL Studies

A much more extensive study of the cross of IHP × ILP generation 70 was reported by Dudley et al. (2007). In this study, phenotypic data were obtained from 500 S₂ lines, derived from the RM7 generation, and their testcrosses to a Mon-

santo company tester. The lines were genotyped for 499 single nucleotide polymorphism (SNP) markers. For the S₂ lines, three environments in each of two years were grown while for the testcrosses three environments were used but only one year. Heritabilities for oil, protein, and starch were all above 0.89 for the S₂ lines per se and around 0.8 for the testcrosses (Table 2). Genetic variance for protein was less than reported by Dudley et al. (2004) for the RM1 of the same cross even though the years and number of progenies were greatly different in the two studies. The number of marker-associated effects (measured by SMA) significant at the 0.01 level for protein and starch was approximately 100 for both the per se and testcross progenies. This number is approximately 20% of the total markers used, a proportion quite similar to that found for the Syn4 in the Dudley et al. (2004) study of the same cross using a much smaller number of markers. The number of markers showing significant effects for oil was approximately half the number showing significant effects for protein and starch. This result differs from those of Dudley et al. (2004) and Goldman et al. (1993, 1994). In the Goldman et al. studies, the number of markers significant for oil was similar to the number for protein and in Dudley et al. (2004) the number significant for oil was only 20% of the number significant for protein.

As with the IHP × ILP cross, a more extensive study of the IHO × ILO cross was reported by Laurie et al. (2004) and Clark et al. (2006). In the IHO × ILO study, 500 S₂ lines, derived from the RM10 generation of the cross of generation 70 of IHO × ILO, and their testcrosses to a Monsanto tester were evaluated using 479 SNP markers. Heritability for oil was approximately 0.93 for both the lines per se and their testcross progeny (Table 2). For protein and starch, heritabilities for

the per se families were about 0.86 and for the testcross families about 0.80. As in other studies cited and as expected from theory (Bernardo, 2002) the genetic variance for testcross progenies was considerably lower than for lines per se (Table 2). The numbers of significant marker-associated effects measured by SMA (50 for the lines per se and 63 for the testcrosses) for oil were similar to the number found in the IHP × ILP cross reported by Dudley et al. (2007). However, the number of significant marker-associated effects for protein was less than half that found in the IHP × ILP cross and the number for starch was slightly more than half that found in the IHP × ILP cross. Thus, approximately 10% of the markers were significant for oil, 8% for protein, and 11% for starch. This compares to approximately 9% significant for oil in the RM5 reported by Willmot et al. (2006) and 3 to 6% significant for protein or starch. The lower proportion of significant markers in the study of Willmot et al. (2006) may have resulted from the increased variability in the Willmot study caused by use of S₃ lines.

Agreement Between IHO × ILO and IHP × ILP

Because of the large number of markers, many of which were closely linked, several markers may be linked to the same QTL. Thus, significant regions were defined as areas where the end of a group of significant markers (based on SMA) was at least 10 cM from the beginning of another region using a Monsanto company composite map. These regions were used to measure agreement in regions identified as significant between the results from the IHO × ILO cross and the IHP × ILP cross (Table 3). Approximately one-third of the SMA regions identified as significant for protein in the IHP × ILP cross were also identified in the IHO × ILO cross (Table 4). Similarly, approximately one-third of the regions identified as significant for oil in the IHO × ILO cross were identified as significant in the IHP × ILP cross.

Table 3. Number of significant single marker analysis regions from the IHO × ILO RM10 cross, the IHP × ILP RM7 cross, and regions in common for traits measured on the per se and testcross families.[†]

Trait	Per se			Testcross		
	IHP × ILP	IHO × ILO	Both	IHP × ILP	IHO × ILO	Both
Protein	27	18	8	31	18	11
Oil	24	21	9	11	14	5
Starch	24	19	15	22	14	8

[†]IHO, Illinois High Oil; ILO, Illinois Low Oil; IHP, Illinois High Protein; ILP, Illinois Low Protein.

Thus divergent selection for protein allowed identification of regions significant for oil not identified in the IHO × ILO cross and divergent selection for oil allowed identification of significant regions for protein not identified in the IHP × ILP cross. For starch, over three-fourths of the regions significant in the per se data for the IHO × ILO cross were also significant in the IHP × ILP cross while over half of the regions significant in the testcross data for IHO × ILO were also significant in the IHP × ILP cross.

Location of QTL

One of the benefits of QTL mapping studies is the ability to locate the QTL affecting a trait to particular chromosomal regions. In general, QTL for oil, protein, and starch were found on all chromosomes. However, certain chromosomal regions consistently had significant, relatively large, effects in most studies. One such region affecting oil concentration is found on chromosome 6. This region was found significant in the RM10 of the IHO × ILO cross, the RM7 of the IHP × ILP cross, the RM1 and RM5 of the IHO × ILO cross, and the Syn0 and Syn4 of the IHP × ILP cross. Likewise a significant region affecting protein concentration was found on chromosome 3 in most studies. In addition to locating QTL affecting particular traits, the biochemical pathway being affected by a segregating QTL can be isolated if gene sequences are used as markers.

Table 4. Number of simple interval mapping regions (on diagonal) significant for protein, oil, and starch in IHP × ILP RM7 and IHO × ILO RM10 and number of regions in common (off diagonal) between traits.

Trait	Cross [†]	Protein		Oil		Starch	
		Per se	Testcross	Per se	Testcross	Per se	Testcross
Protein	IHP × ILP	50	41	34	33	40	36
	IHO × ILO	47	47	31	27	24	26
Oil	IHP × ILP			70	63	40	41
	IHO × ILO			51	54	27	30
Starch	IHP × ILP					60	52
	IHO × ILO					42	42

[†]IHO, Illinois High Oil; ILO, Illinois Low Oil; IHP, Illinois High Protein; ILP, Illinois Low Protein.

Estimates of Number of QTL

Schon et al. (2004) demonstrated that the number of QTL identified in a given cross increased as the number of progeny tested increases. Because of the large and similar numbers of lines and markers used, the IHP × ILP data reported by Dudley et al. (2007) and the IHO × ILO data reported by Clark et al. (2006) were used for estimating numbers of QTL. In the preceding sections, experiments were compared using the number of markers or regions showing significant effects from SMA. This was done because it was about the only way to relate the results of the studies comparing different random mating generations (Dudley et al., 2004; Willmot et al., 2006) with the extensive studies of the advanced random-mated generations (Laurie et al., 2004; Clark et al., 2006; Dudley et al., 2007). However, because many markers are closely linked, particularly in studies using hundreds of markers, comparisons of numbers of significant markers does not reflect the number of QTL. Thus for estimates of numbers of QTL, a measure of the number of significant chromosomal regions was developed. In the studies of IHO × ILO by Clark et al. (2006) and of IHP × ILP by Dudley et al. (2007) regions were defined by simple interval mapping (SIM) where the end of one region was at least 10 cM from the beginning of the next region.

The number of significant regions defined by SIM for the IHP × ILP RM7 and the IHO × ILO RM10 ranged from 41 to 50 for protein, 54 to 70 for oil, and 42 to 60 for starch (Table 4). The numbers for oil are similar to the number (54) estimated in generation 70 using classical quantitative genetic procedures (Dudley, 1977) although the number (122) estimated for protein by Dudley (1977) was somewhat higher than the numbers estimated here. Several points are of interest. First, the number of significant regions for protein was very similar for the two crosses. Second, the number of significant regions in the testcrosses was similar to the number for the lines per se for all traits. For oil, surprisingly, the number of significant regions measured by SIM was larger in the IHP × ILP cross than in the IHO × ILO cross (Table 4). For starch, the number of significant regions for the IHP × ILP cross was larger than for the IHO × ILO cross.

The finding of relatively similar numbers of significant regions in the IHP × ILP cross and the IHO × ILO cross for both protein and oil are puzzling in that divergent selection for protein in IHP × ILP would be expected to have created parents with more loci having divergent alleles for protein than would divergent selection for oil in the IHO × ILO cross. Likewise, divergent selection for oil would be expected to have created divergence for more loci for oil in the IHO × ILO cross than would divergent selection for protein in the IHP × ILP cross. These expectations are supported by the large divergence in means for protein between IHP and ILP and the relatively small difference in means for oil between IHP and ILP (Table 5). Similarly the difference in means for oil between IHO and ILO was large and the difference between those strains for protein was small. The agreement in number of QTL between per se and testcross progenies also is puzzling because the genetic variance in the testcrosses is always considerably less than in the per se progenies (Table 2). However, for all three traits (protein, oil, and starch) the number of significant regions was similar in the per se and testcross progenies (Table 4). In an effort to understand this apparent anomaly, the magnitude of significant effects in the two crosses was compared (Table 2) using SMA data. For oil, the absolute value of significant effects in the IHP × ILP cross was considerably lower than in the IHO × ILO cross. For protein, the absolute value of significant effects in the IHO × ILO cross was lower than in the IHP × ILP cross. In both crosses the absolute value of significant effects is smaller for testcrosses than for lines per se. The question then became how could smaller effects for oil be detected in the IHP × ILP cross than in the IHO × ILO cross and in testcrosses compared to lines per se? A possible answer lies in the fact that the among families within marker genotypes mean square, which is the error term for testing significance of marker-associated effects, includes the genetic variance for the trait being evaluated. For oil, the genetic variance in the IHO × ILO cross is much larger than in the IHP × ILP cross and the average error mean square for testing significance of marker-associated effects is larger (Table 2). Thus, it is possible to detect smaller effects for oil in the IHP × ILP cross. Similarly, the genetic variance for protein is lower in the IHO × ILO cross than in the IHP × ILP cross and thus the error mean square for protein is smaller in the IHO × ILO cross and smaller effects can be detected as significant. Likewise, the genotypic variance in testcrosses is smaller than for lines per se resulting in lower error mean squares for testcrosses than for lines per se.

Another way of viewing this result, is to ask the following questions for both per se and testcross data sets: (i) how many regions in the IHO × ILO oil data sets had effects whose absolute value was greater than the smallest significant effect for oil in the corresponding data set from the IHP × ILP cross and (ii) how many regions in the IHP × ILP protein data sets had effects whose absolute value was greater than the smallest significant effect for protein in the corresponding sets

Table 5. Mean oil and protein concentrations for the generation 70 parents of the extensive quantitative trait loci studies.

Strain [†]	Oil		Protein
	g kg ⁻¹		
IHP	48		266
ILP	31		44
IHO	166		142
ILO	40		118

[†]IHO, Illinois High Oil; ILO, Illinois Low Oil; IHP, Illinois High Protein; ILP, Illinois Low Protein.

from the IHO × ILO cross? Using the SMA data, the smallest significant effect was identified, the numbers of markers determined and those markers separated into regions as defined earlier. Based on these results, if the experiments with oil for IHO × ILO were sufficiently precise to detect effects for oil as small as could be detected in the IHP × ILP cross, two to three times as many regions would have been found significant in the IHO × ILO cross (Table 6). A similar relationship held for protein and the testcross data. Thus the finding of similar numbers of significant regions for oil and protein in the IHP × ILP and IHO × ILO crosses is a function of the precision of the tests used and not necessarily an indication that the same number of QTL were segregating for a trait in both crosses.

These results suggest that neither divergence in means of parents nor magnitude of genetic variance is a good predictor of number of QTL detectable as significant. This agrees with the results of Schon et al. (2004) who found significant numbers of QTL in a cross of two corn inbreds which did not differ markedly for the traits studied. The fact that in IHO × ILO there were QTL effects for oil larger than any found for oil in IHP × ILP suggests that selection for divergence in oil in the IHO × ILO cross had developed closely linked blocks of alleles for high oil in IHO and for low oil in ILO. Larger QTL effects for protein in the IHP × ILP cross suggest that similar blocks of alleles had been developed by divergent selection for protein. Apparently these tightly linked blocks of alleles were not broken by random mating.

Correlated Response

Quantitative trait loci results were used to evaluate relationships among traits by measuring the number of significant SIM regions in common between traits in the IHO × ILO and IHP × ILP crosses (Table 3). In the IHP × ILP cross 40 of 50 SIM regions significant for protein contained significant QTL for starch in the per se progenies and 36 of 41 regions significant for protein in the testcross progenies contained significant QTL for starch (Table 4). Thus, more than 80% of the significant regions for protein were also significant for starch. By contrast, only about one-half of the regions having significant oil QTL had significant QTL for protein. For the IHO × ILO cross, slightly more than half the regions significant for protein in the per se and testcross progenies were significant for starch. For oil, approximately two-thirds of the regions significant for oil in the IHP × ILP cross were also significant for starch whereas in the IHO × ILO cross approximately half of the regions significant for oil were also significant for starch. These results parallel the genetic correlation results.

Table 6. Number of regions with markers significant at 0.01 from single marker analysis or with absolute value of effects greater than minimum significant value in the opposite cross (e.g. IHP × ILP if data are from IHO × ILO).

Cross [†]	Trait	Progeny	No. significant at 0.01	No. greater than minimum
IHO × ILO	Oil	Per se	21	43
IHO × ILO	Oil	Testcross	14	43
IHP × ILP	Protein	Per se	24	41
IHP × ILP	Protein	Testcross	31	41

[†]IHO, Illinois High Oil; ILO, Illinois Low Oil; IHP, Illinois High Protein; ILP, Illinois Low Protein.

QTL Summary

Quantitative trait loci studies agree with classical quantitative genetic results in predicting that oil, protein, and starch concentration in the corn kernel are controlled by large numbers of genes, and that selection for protein will result in lower starch concentration, and little change in oil. However, only the QTL studies can lead to location of the QTL involved and to identification of the nature of the genes involved. Comparison of results from large-scale studies of advanced random-mated generations of IHP × ILP and IHO × ILO demonstrated that neither divergence in means of parents nor magnitude of genetic variance was a good predictor of number of QTL detectable as significant.

Explanation of Long-term Response

The explanation of long-term response in the Illinois selection experiment has been of interest for many years. Dudley (1977) proposed that the presence of large numbers of loci with low frequency of favorable alleles in the original population could account for the progress observed. Goodnight (2004) and Eitan and Soller (2004) suggested epistasis as an important factor. Eitan and Soller (2004) also suggested changes in environment had a major influence on continued response in a selection experiment in chickens. Walsh (2004) suggested mutation was necessary to explain the results. Each of these factors is considered as follows.

1. Large numbers of loci at low gene frequency in the original population. The results presented here, both from “classic” quantitative genetic studies and QTL studies suggest the presence of a large number of segregating QTL affecting oil and protein in the original Burr’s White cultivar. The large number of QTL and estimates of low allelic frequency in the original population, support the hypothesis advanced by Dudley (1977) that the long-continued response to selection could be explained by the presence of a large number of QTL with relatively low gene frequency segregating in the base population. The number of QTL estimated is likely a lower limit for the number segregating

in the original Burr's White because of the small number of parents (5–7) used in making the crosses studied.

2. Release of epistatic variability. When epistasis exists, selection can generate “new” additive genetic variance rather than depleting it as expected under single locus genetic theory (Goodnight, 2004). In simulation studies, epistasis increased the number of generations during which there was a linear response rather than increasing the rate of response (Goodnight, 2004; Liu et al., 2003). Thus Goodnight (2004) proposed epistasis must be considered as a possible cause of long-continued selection response in the Illinois long-term selection experiment. Eitan and Soller (2004) presented the same concept and called it *selection induced genetic variation*. They argued that many QTL exhibit epistatic interactions with other QTL or with genetic background. Because selection changes gene frequency at many loci, the genetic background changes. Quantitative trait loci which are neutral in one genetic background may become major QTL as the background changes. Thus new genetic variability is continually released and may account for long-continued selection response. Data supporting this assumption were presented by Carlborg et al. (2006) who showed in broiler chickens that an apparently minor QTL when placed in alternate homozygous backgrounds of other QTL had a major effect in one background and no effect in the alternate background.

Eitan and Soller (2004) suggested negative heterosis in the cross of a selected population to the original population or an earlier version of the population would provide evidence of the presence of additive × additive epistasis. Crosses of IHO × ILO, IHO × RLO, SHO × ILO, SHO × RLO, and RHO × ILO all showed significant negative heterosis for oil (Table 7; Dudley et al., 1977). Heterosis was positive for the crosses of IHO × RHO and SHO × RHO. Thus genetic background affected heterosis, supporting the hypothesis that additive × additive epistasis was important. Further evidence comes from the Design III study of Moreno-Gonzalez et al. (1975) where crosses of both the F₂

and F₆ of the cross of IHO × ILO back to the parents exhibited negative heterosis for oil (Table 7). This hypothesis is further supported by the presence of significant negative heterosis for protein in the crosses of IHP × RLP and IHP × ILP (Table 8; Dudley et al., 1977). Although dominance for low oil or protein might be considered a cause of the negative heterosis for oil or protein it would not explain the positive heterosis in the crosses of IHO × RHO and SHO × RHO. However both sets of results are consistent with the presence of additive × additive epistasis.

3. Changes in environment. Dudley et al. (1974) suggested part of the continued response in IHP could be due a change in environment because the addition of N fertilizer in generation 53 increased response per generation from 1.4 ± 0.1 to 2.6 ± 0.2 g kg⁻¹ protein per cycle and realized heritability from 0.07 ± 0.02 to 0.12 ± 0.01 in IHP. The increase of available N fertilizer presumably allowed alleles for higher protein to be expressed and selected. Eitan and Soller (2004) made a similar argument to help explain the continued progress over 65 generations from selection for juvenile growth rate in broiler chickens.

4. Mutation. Walsh (2004) argued that mutation was a necessary assumption to explain the results of the long-term selection experiment. Keightley (2004) reviewed selection experiments in inbred lines and concluded mutational variance was important in selection response. However, neither Walsh nor Keightley considered the effects of epistatic interactions on selection response. Goodnight (2004) suggested gene interaction might magnify the effect of beneficial epistatic mutations.

Summary: Explanation of Response

The available data, from both classical quantitative genetics and QTL studies, show that a large number of QTL were segregating in the original Burr's White. Classical quantitative genetic studies suggest that favorable alleles for high oil or protein were present in low frequency in Burr's White. These results support the concept that long-continued response to selection can be explained by a large number of loci segregating in the original population. The presence of negative heterosis in several crosses involving the selection strains for oil and for protein suggests epistatic effects are important. Change in environment affected response to selection for high protein but not necessarily for oil. Although mutation likely played a role in the long-continued progress, it may be subordinate to the other possible explanations and may not be separable from selection induced variation.

Conclusions

Selection works. It is a powerful tool for creating divergent populations from one original population. Large numbers of QTL and, by inference, genes, differentiate IHO and ILO and IHP and ILP for oil, protein, and starch concentration.

Table 7. Midparent heterosis values for oil from crosses among the strains selected for oil.

Parent [†]	IHO	SHO	RHO	RLO	ILO
IHO [‡]		-0.2	1.9*	-2.0*	-3.3*
SHO [‡]			1.6*	-2.2*	-2.4*
RHO [‡]				0.1	-0.8*
RLO [‡]					-0.2
IHO × ILO F ₂ [§]	-2.2				-1.1
IHO × ILO F ₆ [§]	-2.2				-1.4

*Significant at the 0.05 probability level.

[†]IHO, Illinois High Oil; SHO, Switchback High Oil; RHP, Reverse High Oil; RLO, Reverse Low Oil; ILO, Illinois Low Oil.

[‡]Data from Dudley et al. (1977)

[§]Note: test of significance not possible. Data from Moreno-Gonzalez et al. (1975).

This conclusion is supported by both classical quantitative genetic results and the results of QTL studies. Genetic variance and differences in means between parents are not good predictors of the number of QTL which may be found in a particular study. Of more importance are the number of lines evaluated and the number of markers used. From the studies reported here the proportion of marker-associated significant effects is similar in advanced random-mated generations regardless of whether the original cross was IHP × ILP or IHO × ILO and regardless of whether oil, protein, or starch was the trait measured. This result supports the infinitesimal theory of quantitative genetics and suggests that crosses between closely related lines could well lead to identification of a large number of QTL for a given trait if enough progeny and markers were used. This may help explain why progress in narrow-based populations such as those used by corn breeders is so successful.

Taken as a whole, the data support the assumption that progress observed could be the result of a combination of segregation of a large number of genes at low frequency in the original population and selection induced variation resulting from additive × additive epistasis. At the same time, changes in the environment likely affected progress for protein. The possibility that mutation aided selection progress cannot be ruled out.

Acknowledgments

The contributions of the numerous scientists who maintained continuity in the Illinois long-term selection program over more than 100 yr are gratefully acknowledged. A special thanks to R.J. Lambert, a colleague who worked with the author on this experiment for nearly 40 yr, and Don Roberts, the technician on the project for the last 20 yr. Thanks to G.R. Johnson for numerous stimulating discussions as this paper was being written

References

Bernardo, R. 2002. Breeding for quantitative traits in plants. Stemma Press, Woodbury, MN.

Carlborg, Ö., L. Jacobsson, P. Åhgren, P. Siegel, and L. Andersson. 2006. Epistasis and the release of genetic variation during long-term selection. *Nat. Genet.* 38:418–420.

Clark, D.J., J.W. Dudley, T.R. Rocheford, and J.R. LeDeaux. 2006. Genetic analysis of corn kernel composition in the random mated 10 generation of the cross of generations 70 of IHO × ILO. *Crop Sci.* 46:807–819.

Dudley, J.W. 1977. Seventy-six generations of selection for oil and protein percentage in maize. p. 459–473. *In* E. Pollak et al. (ed.) *Proc. Int. Conf. on Quantitative Genetics*, Ames, IA. 16–21 Aug. 1976. Iowa State Univ. Press, Ames.

Dudley, J.W. 1994. Linkage disequilibrium in crosses between Illinois maize strains divergently selected for protein percentage. *Theor. Appl. Genet.* 87:1016–1020.

Dudley, J.W., D. Clark, T.R. Rocheford, and J.R. LeDeaux. 2007. Genetic analysis of corn kernel chemical composition in the random mated 7 generation of the cross of generations 70 of IHP × ILP. *Crop Sci.* 47:45–57.

Dudley, J.W., A. Dijkhuizen, C. Paul, S.T. Coates, and T.R. Rocheford. 2004. Effects of random mating on marker-QTL associations in the cross of the Illinois High Protein × Illinois Low Protein maize strains. *Crop Sci.* 44:1419–1428.

Dudley, J.W., and R.J. Lambert. 1969. Genetic variability after 65 generations of selection in Illinois high oil, low oil, high protein, and low protein strains of *Zea mays* L. *Crop Sci.* 9:179–181.

Table 8. Midparent heterosis values for protein from crosses among the strains selected for protein. Data from Dudley et al. (1977).[†]

	IHP	RHP	RLP	ILP
IHP		–0.3	–2.8*	–4.2*
RHP			0.7	–0.6
RLP				–0.1

*Significant at the 0.05 probability level.

[†]IHP, Illinois High Protein; ILP, Illinois Low Protein; RHP, Reverse High Protein; RLP, Reverse Low Protein.

Dudley, J.W., and R.J. Lambert. 1992. Ninety generations of selection for oil and protein in maize. *Maydica* 37:1–7.

Dudley, J.W., and R.J. Lambert. 2004. 100 generations of selection for oil and protein in corn. *Plant Breed. Rev.* 24:79–110.

Dudley, J.W., R.J. Lambert, and D.E. Alexander. 1974. Seventy generations of selection for oil and protein concentration in the maize kernel. p. 181–212. *In* J.W. Dudley (ed.) *Seventy generations of selection for oil and protein in maize*. CSSA, Madison, WI.

Dudley, J.W., R.J. Lambert, and I.A. de la Roche. 1977. Genetic analysis of crosses among corn strains divergently selected for percent oil and protein. *Crop Sci.* 17:111–117.

Eitan, Y., and M. Soller. 2004. Selection induced genetic variation. p. 154–176. *In* S. Wasser (ed.) *Evolutionary theory and processes: Modern horizons, papers in honour of Eviatar Nevo*. Kluwer Academic Publ., Dordrecht, the Netherlands.

Goldman, I.L., T.R. Rocheford, and J.W. Dudley. 1993. Quantitative trait loci influencing protein and starch concentration in the Illinois long-term selection maize strains. *Theor. Appl. Genet.* 87:217–224.

Goldman, I.L., T.R. Rocheford, and J.W. Dudley. 1994. Molecular markers associated with maize kernel oil concentration in the Illinois High Protein × Illinois Low Protein Cross. *Crop Sci.* 34:908–915.

Goodnight, C.J. 2004. Gene interaction and selection. *Plant Breed. Rev.* 24(part 2):269–291.

Hopkins, C.G. 1899. Improvement in the chemical composition of the corn kernel. p. 205–240. *Ill. Agric. Exp. Stn. Bull.* 55.

Keightley, P.D. 2004. Mutational variation and long-term selection response. *Plant Breed. Rev.* 24(part 1):227–247.

Laurie, C.C., S.D. Chasalow, J.R. Ledeaux, R. McCarroll, D. Bush, B. Haige, C. Lai, D. Clark, T.R. Rocheford, and J.W. Dudley. 2004. The genetic architecture of response to long-term artificial selection for oil concentration in the maize kernel. *Genetics* 168:2141–2155.

Leng, E.R. 1962. Results of long-term selection for chemical composition in maize and their significance in evaluating breeding systems. *Z. Pflanzenzuecht.* 47:67–91.

Liu, P., J. Zhu, X. Lou, and Y. Lu. 2003. A method for marker-assisted selection based on QTLs with epistatic effects. *Genetica* 119:75–86.

Moreno-Gonzalez, J., J.W. Dudley, and R.J. Lambert. 1975. A design III study of linkage disequilibrium for percent oil in maize. *Crop Sci.* 15:840–843.

Schon, C.C., H.F. Utz, S. Groh, B. Truberg, S. Openshaw, and A.E. Melchinger. 2004. Quantitative trait locus mapping based on resampling in a vast maize testcross experiment and its relevance to quantitative genetics for complex traits. *Genetics* 167:485–498.

Smith, L.H. 1908. Ten generations of corn breeding. p. 457–575. *Ill. Agric. Exp. Stn. Bull.* 128.

Walsh, B. 2004. Population and quantitative genetic models of selection limits. *Plant Breed. Rev.* 24:177–225.

Willmot, D.B., J.W. Dudley, T.R. Rocheford, and A. Bari. 2006. Effect of random mating on marker-QTL associations for grain quality traits in the cross of Illinois High Oil × Illinois Low Oil. *Maydica* 51:187–199.

Winter, F.L. 1929. The mean and variability as affected by continuous selection for composition in corn. *J. Agric. Res.* 39:451–476.

Woodworth, C.M., E.R. Leng, and R.W. Jugenheimer. 1952. Fifty generations of selection for oil and protein in corn. *Agron. J.* 44:60–65.