Euphytica **92:** 1–11, 1996.

Genetic basis of the evolution of adaptedness in plants

R.W. Allard
University of California at Davis, Davis, CA 95616, U.S.A.

Key words: adaptedness, epistasis, marker assisted dissection, overdominance, purifying selection

Summary

Adaptedness is both complexly inherited and much affected by environment: consequently the genetic mechanisms that have led to improvements in adaptedness have been difficult to identify and to quantify. Recently it has been shown that 'marker assisted dissection' of adaptedness based on changes in the frequencies of discretely inherited alleles of loci of various kinds (e.g. allozyme, restriction fragment, microsatellite loci) is practicable. I will illustrate marker assisted analysis of the genetic basis of adaptedness with a sample of allozyme data from three species groups, two heavily selfing groups (two wild *Avena* species and barley) and one outcrossing species (corn, maize). The results lead to three main conclusions: (1) that the single most important genetic mechanism in all three species groups was the assembly of favorable epistatic combinations of alleles of different loci by means of recurring cycles of selection, intercrossing superior selects, and inbreeding to near homozygosity leading to stable superior multilocus genotypes adapted to specific habitats; (2) that exploitation of favorable interactions among alleles of the same locus played a significant role in tetraploid *A. barbata* and probably also in single-cross maize hybrids; (3) that purifying selection (elimination of deleterious alleles) played a small role in all three species groups. These results indicate that marker alleles provide applied breeders with effective ways to identify, track, and incorporate regions of chromosomes with favorable effects of adaptedness into improved cultivars.

Introduction

Modern cultivars are much better adapted and much more productive in agricultural environments than their wild ancestors. Adaptedness and productivity are, however, both complexly inherited traits and much affected by environment so that we still know little about what happened genetically during the domestication process. In recent years it has been found that examination of changes in the frequencies of discretely distinguishable alleles and genotypes of morphological, disease resistance, allozyme, restriction fragment, and microsatellite loci that occurred during the domestication process allow deductions concerning the genetic mechanisms that have led to improved adaptedness (e.g. Clegg & Allard, 1972; Allard, 1988, 1990; Perez de la Vega, Garcia & Allard, 1991; Garcia et al., 1991; Allard et al., 1993; Saghai Maroof et al., 1994; Cluster & Allard, 1995). In all cases that have been studied in detail it has been found that population behavior can be explained on the basis of selection acting on the chromosomal segments in which the marker loci reside; thus marker assisted dissection of the genetic basis of adaptedness is feasible and a substantial literature has built up, too large to review in the time allotted. Consequently I will limit discussion to only one type of marker (allozymes) and to three representative species. First I will examine changes in allelic and genotypic frequencies that occurred in the diploid Slender Wild Oat (*Avena hirtula*) and its tetraploid descendant (*A. barbata*) as these wild species evolved over thousands of generations in the Mediterranean Basin. Then I will turn to allelic and genotypic frequency changes that occurred during the domestication and spread of barley (*Hordeum vulgare*) and (*Zea mays*) over hundreds of generations in cultivation. The two *Avena* species and barley are self fertilized (99%) whereas in maize populations the mating system is one of about 90% outcrossing and 10% selfing. The results lead to the conclusion that the single most important genetic

mechanism involved in all three species groups was the development, and the stabilization by inbreeding, of favorable epistatic combinations of alleles of different loci to produce stable multilocus genotypes that provide superior adaptedness in specific environments. A second but less important genetic mechanism was purifying selection (elimination of deleterious alleles). A third genetic mechanism, namely exploitation of favorable interactions among alleles of the same locus (parallel to overdominance) was important in *A. barbata*, a diploidized tetraploid, and to a lesser extent in corn, but overdominance was of little or no importance in populations of diploid *A. hirtula* or barley (no cost effective method of producing F1 hybrid seed had been found in barley).

Avena hirtula (Ah) and *A. barbata (Ab)*

Diploid *Ah* (n = 7) chromosomes) is indigenous to the Mediterranean Basin where it occurs in sparse and more or less disjunct stands. *Ah* is often less robust than *Ab* but the two taxa are usually so similar that they cannot be distinguished on the basis of visible phenotypic characteristics alone. Tetraploid *Ab* (n = 14 chromosomes) is much more widely distributed. It occurs in massive stands throughout the Mediterranean Basin and across the Middle East to Nepal. It has also been a successful colonizer in areas with Mediterranean-like climates throughout the world. Tetraploid *Ab* is clearly much more widely adapted and vigorous than diploid *Ah*.

Ab was introduced to California accidentally from Southwestern Spain about 200 generations ago but *Ah* has not been found in California. In the past half century or so most plants of *Ab* have been found in two major climatic zones in California: (1) a semi-arid zone (250–350 mm annual rainfall) including nearly all of Southern and Central California, and (2) a generally mesic zone (350–1,500 mm annual rainfall) extending > 1,000 km northward in the Pacific coastal ranges from 36° N to Northern Oregon (45° N). Nearly all populations in the xeric zone are monomorphic for two simply inherited variants (glabrous leaf sheaths and dark lemmas) whereas nearly all populations in the mesic zone are monomorphic for hairy leaf sheaths and light lemmas. However mesic patches located within the xeric zone (usually in higher elevation, higher rainfall sites), as well as xeric patches located within the mesic zone (often on steep southwest facing slopes) are nearly always polymorphic for the two pairs of variants. These contrasting ecogeographical patterns suggested that the two morphic pairs might 'mark' and perhaps also contribute to the adaptedness of two 'ecotypes' (Turesson, 1922) that had developed in response to differing selective pressures operating in the xeric and mesic habitats.

In 1930, with guidance from the late W.W. Mackie (University of California, Berkeley) I started experiments to test the above hypothesis. The initial experiments were conducted in several small internally heterogeneous sites (each ca. one hectare in area) located ca. 100 km southeast of San Francisco in an area where the xeric and mesic zones form interfaces with each other. Several small subsites within each site were classified into one of five categories, extending from xeric through three intermediate categories to mesic: assignment to these five categories was based strictly on various physical features of the subsite (e.g. slope, exposure to the sun, edaphic characteristics). Each year during the following decade (1931–1940) > 100 plants from each subsite were classified for the simply inherited leaf-sheath and lemma-color variants and also measured for several quantitative traits such as height, dry weight and number of seeds produced on the primary tiller. The main finding was that the xeric and mesic subsites were nearly always monomorphic for the glabrous-dark or the hairy-light morphs, respectively, but that intermediate subsites were polymorphic in frequencies correlated with degree of xerism. Another main finding was that all of the quantitative traits measured were highly variable within all subsites and none were significantly correlated with degree of xerism. These results suggested that additional simply-inherited 'Mendelizing' markers were likely to be more informative in future studies of ecotypic differentiation than continuously-varying complexly-inherited quantitative characters. However, no additional simply-inherited markers had been found by the time World-War II made it necessary to suspend studies of *Ab*. On return from W.W.II I shifted the ecogenetic studies based on simply inherited 'markers' to crop plants such as barley, beans, sorghum and maize because many simply-inherited markers and highly-isogenic recombinant inbred lines were available in these species.

The critical breakthrough pertaining to simply-inherited markers came in 1960 with the introduction of enzyme electrophoresis to plant genetics (Schwartz, 1960). In short order many technologically advantageous electrophoretically-distinguishable *codominant* allozyme markers had been identified and pressed into

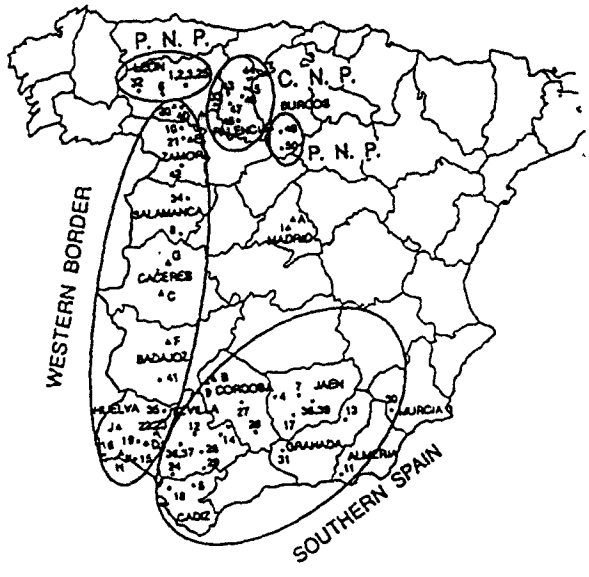

Figure 1. Geographical locations of 10 populations of *A. hirtula* (A-I) and 50 populations of *A. barbata* (1–50).

service in ecogenetic studies in several crop and wild species, including *Ab* and *Ah*. However, the studies carried out in the 1960s, 1970s and 1980s usually focused on only one or two of the several genetic mechanisms that were ultimately found to be jointly responsible for shaping the internal genetic structure, as well as the overall adaptive landscapes of various species. Consequently, I will cite data from only a few recent studies that took into account simultaneously the most important genetic mechanisms that had been identified earlier in more narrowly-focused investigations.

Table 1 gives a sample of allelic and genotypic frequencies for 14 representative codominant allozyme loci in 10 Spanish diploid and 50 Spanish tetraploid populations distributed as shown in Figure 1 (Allard et al., 1993). Note that allelic diversity appears to be greater in the tetraploid than in the diploid e.g. *Ab* has three alleles of *Pgd2* but *Ah* has only one. Counts made on all 14 loci show this is indeed the case: 52 alleles (3.7/locus) were observed in *Ab* but only 38 (2.7/locus) in *Ah*. However only 29 of the 38 alleles of *Ah* were also present in *Ab* indicating that nine alleles of *Ah* were lost during or following the polyploidization process. This raised a question: which alleles were lost and which survived? The answer was clear cut: the nine least-frequent alleles of *Ah*, which ranged in overall frequency from 0.001 to 0.07, were all lost whereas the 29 most-frequent alleles of *Ah*, which ranged in frequency from 0.11 to 1.0, all survived. This raised another question, where did the 23 alleles (52 - 29 = 23) of the tetraploid not present in the diploid come from? Two facts, first that *Ab* is a fully diploidized tetraploid (Hutchinson et al., 1983) and second, that all of the 'new' alleles of *Ab* were found in non-segregating heteroallelic quadriplexes suggest the answer. Locus *Pgd2* (Table 1) serves as a model for the sequence of events that may account for the greater allelic diversity of the tetraploid. Diploid *Ah* is monomorphic for allele *1* of *Pgd2* in Spain (also worldwide); consequently the original quadriplex formed by polyploidization of duplex *11* was almost certainly *11,11*. Theory (Wright, 1937) indicates that the probability is very small that any mutant, even a highly heterotic mutant such as allele *2*, will become established in a small selfing diploid population. In the tetraploid, however, mutational events are equally likely in either of the two genomes and population sizes are typically much larger than in the diploid; hence more than *twice* as many mutations of allele *1* to allele *2* would likely occur in *11,11* quadriplexes than in the *11* duplexes of the diploid. Furthermore, in a diploidized tetraploid such as *Ab*, each such mutation will lead to a *11,21* (or *12,11*) quadriplex heterozygous in one genome, but homozygous in the other genome, and on selfing for four or five generations, at least 1/2 of the progeny are expected to be heterotic true-breeding *11,22* quadriplexes. Consequently it is much more likely that novel favorable mutants would have been incorporated into diploidized tetraploid *Ab* than into diploid *Ah*, in which selfing will rapidly reduce heterozygosity to zero. Subsequent mutations of allele *1* to allele *3* in *11,22* quadriplexes may have led to quadriplex *22,33* of *Pgd2*, which, although less successful than the original *11,11*, or the derivative *11,22* quadriplex, found some sites in which it was able to survive.

Several other patterns of quadriplex formation are illustrated in Table 1. Four other loci of Table 1 (*Mdh3 Acp1, Acp2, Pgd2*) followed a pattern similar to *Pgd2*; each formed a single highly-successful heteroallelic quadriplex (in some cases from two alleles not present in the diploid), as well as one or more less successful heteroallelic or homoallelic quadriplexes. Locus *Mdh1* was unique. It formed a single heteroallelic quadriplex that is now completely monomorphic worldwide. The *intralocus* interaction between pairs of homozygous alleles (*11,22*) of this locus (parallel to overdominance) clearly led to a remarkable level of adaptedness that has thus far withstood all challenges from novel mutant alleles in all environments worldwide. Another pat-

Table 1. Single-locus duplex and quadriplex frequencies in 10 Spanish diploid (*Ah*) populations and 50 Spanish tetraploid (*Ab*) distributed as shown in Figure 1. (Adapted from Allard et al., 1993)

Locus	Diploid			Tetraploid		
	Duplex	f	No. sites in which fixed	Quadriplex	f	No. sites in which fixed
Pgd2[a]	*11*	1.00	10/10	*11,11*	0.12	1/50
				11,22	0.84	24/50
				22,33	*0.04*	0/50
					1.00	
Mdh1	*11*	1.00	10/10	*11,22*	1.00	50/50
Est1[b]	*11*	0.07	0/10	*33,33*	0.15	1/50
	22	0.24	0/10	*55,55*	0.14	0/50
	55	0.44	2/10	*11,33*	0.04	0/50
	66	0.03	0/10	*22,33*	0.19	2/50
	77	*0.19*	0/10	*22,55*	0.15	1/50
		0.97		*33,55*	0.04	0/50
				33,66	0.03	0/50
				55,77	*0.20*	4/50
					0.94	
Pgm1[c]	*11*	1.00	10/10	*11,11*	1.00	50/50

[a] Loci *Pgd2*, *Mdh3*, *Acp1*, *Acp2* and *Pgd1* have similar patterns.
[b] Loci *Est1*, *Lap1*, *Prx1* and *Pgi* have similar patterns.
[c] Loci *Pgm1*, *Got2*, *Mdh2* and *Got1* have identical patterns.
[f] Estimates based on 754 diploid plants (1,508 alleles) and 4,751 tetraploid plants (19,004 alleles).

tern of quadriplex formation is that of four loci (*Est1*, *Lap1*, *Prx1*, *Pgi1*) that are highly polymorphic in the diploid. Each of these loci formed several homoallelic and several heteroallelic quadriplexes. Still another pattern of quadriplex formation is that of loci *Pgm1*, *Got2*, *Mdh2*, and *Got1*. All of these loci are monomorphic, or very nearly so, in Spain (also worldwide) for a single allele in the diploid and also for the same allele in the tetraploid. Evidently these alleles code for some essential function such that they confer adaptedness in homoallelic state superior to that conferred by any other alleles, whether homoallelic or heteroallelic, that have arisen in either *Ah* or *Ab*. Overall the net effect of polyploidization, followed by diploidization, was a great increase in allelic diversity in the tetraploid. For the majority of loci the increases in allelic diversity provided not only opportunities for favorable intralocus interactions (parallel to overdominance) but also greatly increased opportunities for exploitation of favorable epistatic interactions among alleles of different loci. As an example, consider a diploid plant genotypically *11* for one locus (say *Lap1*) and *55* for another locus (say *Est1*). Heterozygotes are so rare in the diploid that exploitation of possible overdominant intraallelic interactions in *15* heterozygotes seems quite unlikely: also only a single epistatic interaction (*11* × *55*) is possible. However, in a tetraploid population that includes individuals genotypically *22,33* for *Lap1* and *55,77* for *Est1*, six interactions are possible, two intralocus (overdominant) interactions, *22* × *33* for *Lap1* and *55* × *77* for *Est1*, as well as four pairwise interlocus epistatic interactions, *22* × *55*, *22* × *77*, *33* × *55 and 33* × *77*. All of these combinations are either fixed or nearly fixed in all seven populations of the cold Central Northern Plateau of Spain, which suggests that these particular interactions contribute to adaptedness in the habitat of that region. Many other non-random associations of particular duplexes and quadriplexes with particular regions and also with specific sites within regions were also observed, which suggests that specific duplexes and quadriplexes confer superior adaptedness in particular regions as well as in specific sites within regions.

Many different two-locus, three-locus, up to 14-locus interactions are also possible in both the diploid and the tetraploid. Patterns in the diploid are, however, much less complex than in the tetraploid. Consequently I now turn to the 14-locus genotypes observed in

Table 2. Most frequent 14-locus genotypes in 10 diploid sites. (Adapted from Allard et al., 1993)

Site	Locus											N[b]	f[c]
	Pgm1[a]	*Got1*	*Mdh2*	*Mdh3*	*Acp2*	*Pgd1*	*Pgi1*	*Acp1*	*Prx1*	*Lap1*	*Est1*		
1	*11*	*11*	*11*	*22*	*11*	*11*	*11*	*55*	*55*	*11*	*55*	1	1.00
2	*11*	*11*	*11*	*66*	*11*	*11*	*44*	*11*	*22*	*22*	*55*	1	1.00
3	*11*	*11*	*11*	*22*	*22*	*11*	*22*	*11*	*11*	*11*	77^P	2	0.93
4	*11*	22^P	*11*	*22*	*11*	*11*	11^P	22^P	*22*	44^P	55^P	8	0.71
5	*11*	*11*	*11*	*22*	*11*	11^P	44^P	*11*	*22*	22^P	55^P	8	0.51
6	*11*	*11*	*11*	*22*	*11*	55^P	*11*	55^P	44^P	44^P	11^P	13	0.45
7	*11*	*11*	*11*	*22*	*11*	22^P	44^P	11^P	22^P	22^P	22^P	20	0.26
8	*11*	*11*	*11*	22^P	11^P	11^P	11^P	22^P	11^P	77^P	77^P	17	0.21
9	*11*	*11*	*11*	*22*	55^P	11^P	22^P	22^P	22^P	34	22^P	34	0.17
10	*11*	*11*	11^P	*22*	22^P	22^P	22^P	11^P	11^P	11^P	55^P	36	0.12

[a] Also *Got2, Mdh1, Pgd2.*
[b] Number of 14-locus genotypes observed at each site.
[c] Frequency within site of most frequent 14-locus genotypes.
[P] Site polymorphic for 14-locus genotypes due to polymorphism at locus indicated.

the diploid to give a feel for some of the two-locus and higher-order interactions that have been successful. In total 107 different 14-locus genotypes were observed in the 10 Spanish populations of *Ah.* The main features of both within-population and between-population differentiation can be deduced from Table 2, which gives the frequencies of the single most-frequent 14-locus genotype, and the total number of 14-locus genotypes, at each site. The array of 14-locus genotypes present at each site differed from that at each other site. Only one 14-locus genotype was present in Sites 1 and 2; the 14-locus genotype at these two sites was identical at nine loci but differed at five loci (*Mdh3, Pgi1, Acp1, Prx1, Lap1*). It is possible that differences in adaptedness between these two genotypes may have been due entirely to the main effects of these five loci but the very large number of different patterns in which the alleles of the 14-locus genotypes occur in the diploid suggest that two-locus, three-locus and higher-order interactions were also involved. In fact discrete log-linear analyses showed that many multilocus interactions are highly significant statistically but others are nonsignificant (Perez de la Vega et al., 1991; Garcia et al., 1991). Within-population diversity varied from none for Sites 1 and 2, to small for Site 3 (only two 14-locus genotypes present), to substantial for Sites 4 and 5 (eight different 14-locus genotypes present), to high for the five remaining sites (13 to 36 different 14-locus genotypes present). Hence opportunities for epistatic interactions were much greater in Sites 3–10 than in Sites 1 and 2. Population sizes were large at each site so that the effects of genetic drift are expected to be small.

Also frequent seed exchange occur between sites (due largely to agricultural activities). Such exchanges are expected to limit genetic differentiation among sites: hence it seems likely that the substantial genetic differences that occurred among sites were due, not to genetic drift, but rather to selection sorting out the single best-adapted 14-locus genotype in the two monomorphic sites, as well as the particular mixes of 14-locus genotypes observed in the polymorphic sites. Rapid changes in the frequencies of 14-locus genotypes are known to occur from generation to generation within populations, particularly in highly polymorphic sites, but the production of new 14-locus genotypes is almost certainly a slow process. The reasons are as follows. The 99% of selfing that occurs in *Ah* (also *Ab*) forces all loci to near homozygosity with the result that *no* recombination takes place generation after generation between the great majority of loci, even loci located on different chromosomes. At the same time the ca. 1% of outcrossing that occurs leads to short bursts of segregation and recombination that produce novel genotypes, some of which may be superior to existing genotypes. If an outcross plant is heterozygous for n loci, half of its selfed progeny are expected to be heterozygous for these loci in the next generation. Thereafter selfing will reduce heterozygosity by 1/2 in each succeeding generation; thus the series for heterozygosity within each lineage descended from any single outcross individual is 1/2, 1/4, 1/8, 1/16 so that > 90% of the recombination occurs in the first four generations following an outcross. Assuming no fitness differences and large numbers of individuals within lineages, each lineage

is soon expected to approach an equilibrium featuring 2^n equally frequent homozygous lines. However it is likely that selection will favor the more fit genotypes and eliminate the less fit, and also that many genotypes will drift out of each lineage. Regardless, the survivors will quickly be driven to homozygosity by the selfing and all loci within each lineage, whether located on the same or different chromosomes, will thereafter behave as if they are tightly linked with recombination value ca. 0.01 (Allard, 1975). Thus any two loci, (say a and b), no matter where they are located within the genome, will behave as a single functional locus c, and they will manifest some degree of pseudo-dominance up to pseudo-*over*dominance whenever $c_1 = a_1b_2$, $c_2 = a_2b_1$, and $c_1c_1 < c_1c_2 > c_2c_2$. Selfing is thus a simple and highly effective way of exploiting all favorable epistatic interactions among alleles of different loci, whether the loci are physically linked on the same chromosome or not. The 1% of outcrossing provides a small but steady supply of novel multilocus genotypes, some of which may be superior, whereas the 99% of selfing restricts recombination sufficiently to protect previously existing, as well as newly arisen favorable epistatic combinations of alleles, from breakup due to segregation. This raises another question: at what rate of outcrossing does the resulting segregation destroy existing population structure? Experiments with barley show that increasing outcrossing rates artificially from the normal 1% to as little as 5%, even for a single generation, leads to dramatic alterations in population structure accompanied by large decreases in seed yields that persist for many generations after selfing is allowed to resume. Evidently even low outcrossing rates can lead to 'outbreeding depression' in populations of heavy selfers.

Turning to *Ab*, the number of 14-locus genotypes observed in the 50 Spanish populations exceeded 440, more than four-fold as many as in *Ah*. Fourteen-locus genetic structure was much more complex than in *Ah* and also appeared to be more closely attuned to specific environments. Consequently I will limit discussion to a single point that emerged more clearly from the *Ab* data than the *Ah* data, namely that very close associations developed between distinctive habitats and certain single 14-locus genotypes, as well as between other specific habitats and particular polymorphic mixtures of 14-locus genotypes. One set of such associations were those with the habitats of the cold, high elevation North Central Spanish Plateau (CNP, Figure 1). A specific 14-locus genotype (Table 3) was most frequent in all seven sites of that region (Sites 43–48, 33); this particular 14-locus genotype was monomorphic (f = 1.00) in Site 43 and nearly monomorphic (f = 0.99) in Site 45, the two highest and coldest sites of the region. Two things happened in the five lower elevation sites of the CNP: (1) the frequency of the 'cold-tolerant' 14-locus genotype fell off (as in Site 46) and/or (2) several additional loci became polymorphic (as in Site 33). The 'cold tolerant' genotype was entirely absent in still lower and less cold areas peripheral to the CNP. In Site 1 (PNP, Figure 1) it was replaced by a 14-locus genotype that differed at only one locus (quadriplex *11,44* of *Acp2* was replaced by quadriplex *33,44*) and in Site 3 in which quadriplex *11,11* of *Prx1* was replaced by quadriplex *11,22*. At the same time these two sites also became polymorphic for several other loci leading to substantial changes in the patterns of multilocus epistatic interactions. In brief, progressively larger environmental changes were accompanied by progressively greater restructuring of the 14-locus genetic makeup of the populations. Evidently natural selection sorted out the single 14-locus genotype that is best adapted in the coldest areas and integrated many different 14-locus genotypes into complex entities that provide superior adaptedness in the less cold environments of lower elevation sites. However among the 50 Spanish populations of *Ab* only four were monomorphic for a single 14-locus genotype whereas more than 20 different 14-locus genotypes were present in a great majority of the populations (Allard et al., 1993). This indicates that superior adaptedness in *Ab*, as in all other heavily selfing species that have been studied in detail, is nearly always associated with substantial genetic diversity featuring complex mixtures of highly-fit non-segregating multilocus genotypes. Heavily selfing populations are thus usually highly variable genetically: they clearly are not stuck in evolutionary 'blind alleys' as has frequently been stated. In fact their genetic system is magnificently equipped to develop and to incorporate novel superior genotypes while at the same time preserving the mix of homozygous nonsegregating genotypes that has evolved over time in response to fluctuating selective pressures exerted by the specific environment they occupy.

Table 3 also gives the 14-locus genotypes of *Ab* in four of seven habitats for which genotype-habitat associations have been established in California. Note that the present day ancestral and colonial populations are closely similar in allelic composition when compared on a locus-by-locus basis: the predominant alleles of Spain are also predominant in the colonial populations. However, the ancestral and colonial populations differ

Table 3. Single most-frequent 14-locus genotypes of *Ab* in the Central Northern Plateau (CNP) and the Peripheral Northern Plateau (PNP) of Spain and in four habitats in California. (Adapted from Allard et al., 1993)

Region and site	Locus *Pgm1*[a]	*Mdh1*	*Pgd2*	*Mdh3*	*Acp2*	*Pgd1*	*Acp1*	*Prx1*	*Lap1*	*Est1*	f[b]
CNP-43	*11,11*	*11,22*	*11,22*	*11,22*	*11,44*	*11,33*	*11,22*	*11,11*	*22,33*	*55,77*	1.00
CNP-45	*11,11*	*11,22*	*11,22*	*11,22*	*11,44*	*11,33*^{P}	*11,22*	*11,11*^{P}	*22,33*	*55,77*	0.99
CNP-46	*11,11*	*11,22*	*11,22*	*11,22*	*11,44*^{P}	*11,33*^{P}	*11,22*	*11,11*	*22,33*	*55,77*^{P}	0.68
CNP-33	*11,11*	*11,22*	*11,22*^{P}	*11,22*^{P}	*11,44*^{P}	*11,33*^{P}	*11,22*	*11,11*^{P}	*22,33*^{P}	*55,77*^{P}	0.91
PNP-1	*11,11*	*11,22*	*11,22*^{P}	*11,22*^{P}	*33,44*^{P}	*11,33*^{P}	*11,22*^{P}	*11,11*^{P}	*22,33*^{P}	*55,77*^{P}	0.48
PNP-3	*11,11*	*11,22*	11,22^P	*11,22*^{P}	*11,44*^{P}	*11,33*^{P}	*11,11*^{P}	*11,22*^{P}	*22,33*	55,77^P	0.33 rDNA
CA-M[c]	*11,11*	*11,22*	*11,22*	*11,22*	*33,44*	*22,22*	*11,33*	*11,22*	*11,33*	*55,55*	*15,9, 8, 7*
CA-X[c]	*11,11*	*11,22*	*11,22*	*11,22*	*33,44*	*11,11*	*11,22*	*11,33*	*11,22*	*33,55*	*13, 10, 8, 7*
CA-H[c]	*11,11*	*11,22*	*11,22*	*11,22*	*33,44*	*11,22*	*11,22*	*11,33*	*11,22*	*33,35*	*10, 8,7*
CA-JR[c]	*11,11*	*11,22*	*11,22*	*11,22*	*33,44*	*11,22*	*11,22*	*11,22*	*11,22*	*33,55*	*12, 8, 7*

[a] Also *Got2, Mdh3, Got1, Pgi1.*
[b] Frequency within site of single most-frequent 14-locus genotype.
[P] Locus polymorphic within site.
[c] Polymorphic in areas where M, X, H and/or JR habitats form interfaces with each other.

in genotypic configurations in two main ways. First, the predominant alleles of Spain often occur in different combinations within quadriplexes in California and second and more striking, the 14-locus genotypes of California are entirely different from those of Spain. Another difference, not documented in Table 3, is that the *rare* alleles of Spain were replaced in California by a nearly entirely different set of *rare* alleles that evidently arose recently as a result of mutations that occurred within the colonial Californian populations. Rare alleles have little effect on adaptedness; consequently it appears that changes in allelic frequency, which are often considered to be the elemental process of evolution, played only a minor role in the adaptive changes that occurred in the colonial populations. Instead, a different process, namely reorganization of the *predominant* alleles of the ancestral populations into novel combinations adapted to specific habitats in California, played the major role. This dramatic reorganization obviously occurred quickly because *Ab* has been widespread in California for no more than 150 generations. The process of adaptation clearly continues up to the present; many populations are highly polymorphic in California, especially in areas where different habitats adjoin one another and frequent hybridizations occur among genetically different ecotypes.

Table 4. Numbers of alleles of allozyme and rDNA loci in 26 collections of wild barley and from cultivated barleys in three stages of domestication

	No. of loci	No. of alleles	Mean no. alleles/locus	Relative no. alleles/locus
[a] Wild barley	20	103	5.15	100
[b] Landraces	20	55	2.75	53
[c] CCII F_7–F_9	26	41	1.58	31
CCII F_{53}	26	37	1.42	28
[c] CCXXI F_4	25	42	1.68	33
CCXXI F_{22}	25	38	1.52	30
[d] Californian cultivars	25	36	1.44	28

[a] From 26 Middle Eastern collections.
[b] Eighteen landraces from the Middle East.
[c] Composite Crosses II and XXI, respectively, were synthesized from a worldwide sample of 28 early 20th Century cultivars and from 6,200 accessions of the USDA world barley collection. Both populations were grown annually in California in plots sufficiently large to avoid genetic drift, harvested in bulk without conscious selection and the next generation was sown from a random sample of seeds from the previous harvest.
[d] Nine modern Californian cultivars adapted in different eco-geographical regions.

Wild and cultivated barley

Table 4 gives the numbers of alleles observed in wild barley (*Hordeum vulgare* ssp. *spontaneum*), and in cultivated barley (*H. vulgare* ssp. *vulgare*) in four distinct

Table 5. Frequency of alleles of 23 allozyme loci in 94 Mexican landrace collections of maize, in the 30 most popular inbreds, and in the six inbreds most widely used in elite single crosses. (Adapted from Doebley et al., 1985)

Allelic frequency in Mexican races	% of Mexican collections in which observed	No. of alleles in frequency classes	No. surviving alleles into: Set of 30 inbreds	Set of six inbreds
0.99–0.78	100%	18	18 (100%)	18 (100%)
0.68–0.21	85%	10	10 (100%)	6 (60%)
0.19–0.05	36%	13	10 (77%)	5 (38%)
0.05–0.01	12%	25	9 (36%)	4 (16%)
< 0.01	2%	97	2 (1%)	1 (1%)

[a] Reid heterotic group: B73, B37, A632; Lancaster heterotic group C103, Oh43, Mo17.

stages of domestication. In the transition from wild barley to so-called primitive Middle Eastern landraces numbers of alleles/locus *decreased* by about one half. However progressively smaller decreases in numbers of alleles/locus occurred in the transitions from Middle Eastern landraces, to early 20th Century cultivars, to the elite Californian barley cultivars, all of which are closely similar in allelic composition. It became apparent early in studies of the Mendelian inheritance of alleles of the loci of Table 4 that patterns of allelic frequency change in segregating families are good predictors of the probable adaptive values of different alleles of most loci. In such Mendelian studies, observed ratios in F_2 families nearly always deviated significantly from expected 3:1 or 1:2:1 ratios; frequent (predominant) alleles were nearly always in excess and infrequent alleles were in deficiency. Thus, even under the conditions of little or no plant-to-plant competition under which the Mendelian studies were conducted, rare and infrequent alleles nearly always behaved as subvitals. This was also the case in large populations (such as *CCII*) in which interplant competition was intense and genetic drift was negligible. However alleles that were present in intermediate frequencies in the wild, as well as in landraces and early 20th Century varieties, usually followed a quite different pattern of allelic frequency change in large experimental populations. Increases (or decreases) in the frequencies of such alleles were usually slow (0.01 per generation). Also the superior (inferior) adaptive properties of allozyme alleles that ultimately became predominant (or infrequent) were rarely if ever detectable in visual comparisons of pairs of highly isogenic recombinant inbred lines, one of which carried one allele and the other an alternative allele of the locus under investigation. Furthermore, seed yields of such isogenic lines measured in head-to-head small-plot comparisons were usually not significantly different. Although differences between such isogenic lines were clearly subtle, natural selection clearly recognized superior vs. inferior adaptive properties of individual alleles or the very short chromosome segments in which these alleles reside.

Studies of dynamic change have also provided clues regarding another feature of the evolution of cultivated barleys, namely, how populations that are so similar in allelic composition as the landraces, the early 20th Century varieties and more particularly the modern elite cultivars can differ so widely in adaptedness and performance. Comparisons of multilocus genetic structure have shown that populations in different evolutionary stages differ relatively little, either in allelic composition or allelic frequencies, but that they differ widely in the *combinations* in which surviving alleles occur in the multilocus genotype of each cultivar. I will now illustrate this point more precisely with data from corn.

Maize

It is widely accepted that Southwestern Mexico was the cradle of domestication of maize and that North American maize was derived from Mexican races. Table 5 gives the frequencies of 163 alleles of 23 allozyme loci in 94 collections representing 34 different Mexican races. Among these 163 alleles 18 (11%) were present in all 94 Mexican collections at overall high frequencies from 0.78 to 0.99. *All* of these ubiquitous and frequent alleles survived into the 30 most-popular U.S. public inbred lines and *all* of them also survived into the six inbred lines that were used as parents of

virtually all U.S. public single-cross hybrids. The situation was the opposite for the 97 alleles (60%) that were present in frequencies < 0.01. These alleles were rarely present in more than one or two Mexican races; also only two among these 97 rare alleles survived into the 30 most popular inbreds and only one survived into the six elite inbreds. The remaining 48 (29%) among the total of 163 alleles were present in Mexican races (also in Bolivian races) in intermediate frequencies. All of the 10 alleles present in relatively high frequencies (0.21 to 0.68) and 10 of the 13 alleles present in moderate frequencies (0.05 to 0.19) survived into the 30 most-popular inbreds; however only about half of these alleles survived into the set of six most-widely-used inbred lines and none were fixed in these six elite lines. Thus, in maize, as in *Avena* and in barley, frequent (predominant) alleles contribute to adaptedness in many habitats and survive many cycles of selection in cultivation whereas alleles that are present in intermediate frequencies overall appear to be useful in some environments but not in others; rare alleles of allozyme loci appear to be of little value anywhere. The data for precisely identifiable alleles of maize as well as barley, therefore support the notion that plant breeding has led to a reduction in allelic diversity; however the data also indicate that the reduction in allelic diversity that occurred is due largely to purifying selection rather than to 'erosion' of useful genetic variation.

The data of Table 6 show that substantial reorganization occurred at the single-locus level during the breeding of the elite public single-cross hybrids and that the reorganization that occurred at single loci ran nearly entirely counter to expectations based on the widely-accepted hypothesis that heterosis results from advantage of heterozygotes over homozygotes. Under the overdominance hypothesis it is expected that the two parental inbreds of high performing single-cross hybrids will have diverged maximally in allelic frequency so that many loci will be heterozygous in single crosses. However, among the 23 loci of Table 6 maximum divergence occurred only for locus *Glu1*; all three inbreds of the Reid heterotic group carry allele *1* and all three inbreds of the Lancaster heterotic group carry allele *2* so that the Lancaster × Reid single crosses are all *12* heterozygotes for *Glu1*. In sharp contrast, both the Lancaster and Reid inbreds carry the same allele for 14 of the 23 loci (61%) so that all Lancaster × Reid single crosses are *11* homozygotes for these 14 loci. The six most-widely-used inbreds were sometimes fixed for different alleles of the eight remaining loci; consequently some Lancaster × Reid single crosses were homozygous and some were heterozygous for these loci. Overall the data of Table 6 show that the great majority of loci (79%) are homozygous and only 21% were heterozygous in the most-widely-grown public single-cross hybrids. The high proportion of homozygotes and the low proportion of heterozygotes at the single-locus level cast doubts about the importance of overdominance in promoting high performance.

It is well known that F_2 seed produced on F_1 single-cross hybrids produce about 10% to 20% lower yields than the F_1 single-cross hybrids themselves. All plants of an F_1 single-cross population are genetically identical so that saving seed produced on open-pollinated single-cross plants is equivalent to a single generation of selfing. Hence producing seed from single-cross populations reduces heterozygosity by half at each heterozygous locus. Is this reduction in heterozygosity the cause of the large decrease in seed yield? Probably not, because it seems unlikely that reducing the frequency of heterozygotes from 1.0 to 0.5 at only 21% of loci would cause such dramatic decreases in performance.

The data of Table 6 show further that extraordinary reorganization took place at the 23-locus level during the breeding of the most-widely-grown public single crosses. Millions of different 23-locus allozyme genotypes are present in the Mexican races, in U.S. open-pollinated varieties of maize, and also in the double-cross hybrids that can be produced from the 30 most popular inbreds, but the number of 23-locus allozyme genotypes had been reduced dramatically to only four in the most-widely-grown public single crosses. Apparently repeated testing for favorable interactions among alleles from different gene pools, such as the Lancaster and Reid gene pools, combined with periodic stabilizing of favorable combinations by inbreeding to homozygosity, reduced the milions of 23-locus genotypes in the Mexican races, open-pollinated varieties and double-cross hybrids to only four 23-locus genotypes observed in the most-widely-grown public single-cross hybrids. Note, however, that very large numbers of pairwise, three-locus, four-locus and higher-order epistatic interactions are possible among the 23 loci. Do favorable interlocus epistatic interactions in fact play a major role in the superior performance of the elite public single crosses? About one in five of the allozyme loci (21%) of the public single crosses is heterozygous and hence it seems likely that segregation and recombination among such loci might quickly dismantle the single 23-locus monomorphic genotype of each of the elite public single-cross populations. Examination of the allozyme genotypes of pop-

Table 6. Twenty-three locus allozyme genotypes in F_1 single-crosses among the six most-widely used-public-inbred lines of maize in North America[a]. (Adapted from Goodman & Stuber, 1980)

Fourteen[b] monomorphic loci	Nine polymorphic loci[c]									Twenty-three locus genotypes
	1	2	3	4	5	6	7	8	9	
11	*12*	*11*	*11*	*12*	*12*	*11*	*11*	*11*	*11*	I
11	*12*	*11*	*11*	*11*	*11*	*12*	*12*	*11*	*33*	II
11	*12*	*12*	*12*	*11*	*11*	*11*	*12*	*12*	*13*	III
11	*12*	*12*	*11*	*11*	*11*	*12*	*12*	*12*	*12*	IV

[a] Reid heterotic group: B73, B37, A632; Lancaster heterotic group C103, Oh43, Mo17.
[b] *Adh1, Cat3, Ep, Got1, Got2, Got3, Idh1, Mdh1, Mdh2, Mdh4, Mdh5, Me, Mmm, Pgm1.*
[c] 1 *Glu1*, 2 *Idh2*, 3 *Pgd1*, 4 *Pgd1*, 5 *Pgd2*, 6 *Pgm2*, 7 *Phi1*, 8 *Est8*, 9 *Acp1*.

ulations derived from seed produced on single crosses showed that many different 23-locus genotypes were present in such F_2 populations. In fact when a sample of 50 or more F_2 progeny of the elite F_1 single crosses was examined no single 23-locus genotype was observed more than once in any F_2 progeny; also in no case did the unique 23-locus genotype of the F_1 single cross appear in its F_2 progeny. Thus in a single generation and recombination apparently completely dismantled the single 23-locus genotype of the most-elite public single-cross hybrids and replaced this genotype with a complex mix of less fit genotypes.

Implications for plant breeders

The main messages for plant breeders that emerge from the above results can be summarized in terms of three generalizations. Generalization 1. The most useful genetic resources are modern elite cultivars and their close relatives, especially materials that are adapted in the local environment or closely similar environments. Natural selection in combination with breeder-directed selection in farmers' fields and breeders' nurseries practiced over large numbers of generations clearly increased the frequencies of favorable alleles and of favorable lower-order combinations of alleles, and at the same time decreased the frequencies of less favorable alleles and less favorable lower-order combinations of alleles. But having many favorable alleles and favorable lower-order combinations present is not enough. Higher-order combinations are also very important. This leads to Generalization 2. Once favorable multiallelic combinations have been developed for a given habitat it is important that such combinations be preserved and enhanced. The most effective way to preserve and enhance favorable combinations is to cross elite materials with elite relatives. Also selfing such crosses is preferable to sib crossing because the wider the cross the greater the chance that segregation will dismantle favorable multilocus combinations. However need may arise ultimately to introgress exotic alleles into the elite materials. The difficulties of doing this depend on the number and the heritability of the alleles to be introgressed. This leads to Generalization 3. Discretely inherited markers often provide breeders with effective ways of identifying, tracking and incorporating regions of chromosomes with favorable effects on adaptedness into elite materials. I will illustrate this with two examples, one relatively simple and one more complex.

The relatively simple example involves a photoperiod-sensitive male-sterile (*pms*) rice plant that was found as a spontaneous mutant in a Chinese cultivar in 1973. Early studies indicated that this mutant could be used to propagate itself under short-day conditions and also to produce F_1 hybrid seeds by interplanting it with normal fertile lines under long-day conditions. Thus *pms* rice appeared to offer great opportunities for replacing the widely used 'three-line' (male-sterile, maintainer, restorer) system with a more efficient 'two-line' system. However, it turned out to be less straightforward than first thought because it was found that, in addition to photoperiod, temperature plays an important role in the fertility of many rice genotypes. The solution turned out to be marker-assisted-dissection of fertility, which revealed that nearly all of the *pms* effect, including the temperature effect, was governed by two genetic loci, one very tightly linked to an RFLP marker on chromosome 7 and another very tightly linked to another RFLP marker on Chromosome 3 (Zhang et al., 1994). This made it possible, in developing *pms* lines, to identify and track the *pms* effect precisely so that laborious, costly and

time-consuming measurements of fertility were unnecessary. Chinese rice breeders anticipate that many different elite two-line hybrids adapted in different habitats will soon be available for commercial production.

The second example involves marker dissection of adaptedness in barley Composite Cross II in California. CCII was synthesized in 1928 by compositing hybrid seeds from all possible pairwise combinations (378) among carefully chosen barley varieties representing all major barley growing regions of the world. All of the 30 or so markers ultimately studied in CCII turned out to have substantial effects on adaptedness. Interactions between genotypes and environments were, however, very complex. Early studies at Davis were especially confusing – some markers steadily increased and others steadily decreased in frequency while others varied in behavior, sometimes increasing, sometimes decreasing but sometimes not changing in frequency. The reasons emerged when CCII was grown in areas of California where rainfall was either consistently limiting or rarely limiting. In such areas nearly all markers behaved quite consistently; the inconsistency at Davis stemmed from the fact that about one year out of three featured either severe moisture stress, limited moisture stress, or no moisture stress. Thus, at Davis, CCII chased a different multilocus adaptive peak in about one year out of three, whereas in several other locations CCII chased the same adaptive peak consistently so that it was soon possible to identify the multilocus marker genotype appropriate to each habitat. The interactions among marker loci were complicated but far less difficult to identify and quantify than the interactions of quantitatively inherited characters with each other and with environmental variables. I anticipate that marker assisted breeding will provide helpful guidance in developing varieties with superior adaptedness in the several diverse habitats in which barley is grown in California.

Acknowledgement

Dedicated to Göte Turesson on the 100th anniversary of his birth.

References

Allard, R.W., 1975. The mating system and microevolution. Genetics 79: 115–126.

Allard, R.W., 1988. Genetic changes associated with the evolution of adaptedness in cultivated plants and their wild progenitors. Jour Hered 79: 225–238.

Allard, R.W., 1990. The genetics of host-pathogen coevolution: Implications for genetic resource conservation. Jour Hered 81: 1–6.

Allard, R.W., P. García, L.E. Sáenz de Miera & M. Pérez de la Vega, 1993. Evolution of multilocus genetic structure in *Avena hirtula* and *Avena barbata*. Genetics 135: 1125–1139.

Clegg, M.T. & R.W. Allard, 1972. Patterns of genetic differentiation in the slender wild oat species *Avena barbata*. Proc Natl Acad Sci USA 69: 1820–1824.

Cluster, P.D. & R.W. Allard, 1995. Evolution of ribosomal (DNA) genetic structure in colonial Californian populations of *Avena barbata*. Genetics 139: 941–954.

Doebley, J.F., M.M. Goodman & C.W. Stuber, 1985. Isozyme variation in the races of corn from Mexico. Amer J Bot 72: 629–639.

García, P., M.I. Morris, L.E. Sáenz de Miera, R.W. Allard, M. Pérez de la Vega & G. Ladizinsky, 1991. Genetic diversity and adaptedness in tetraploid *Avena barbata* and its diploid ancestors *A. hirtula* and *A. wiestii*. Proc Natl Acad Sci USA 88: 1207–1211.

Goodman, M.M. & C.W. Stuber, 1980. Genetic identification of lines and crosses using isoenzyme electrophoresis. Annual Corn Sorghum Res Conf Proc 35: 10–31.

Hutchinson, E.S., S.C. Price, A.L. Kahler, M.I. Morris & R.W. Allard, 1983. An experimental verification of segregation theory in a diploidized tetraploid: esterase loci in *Avena barbata*. J Heredity 74: 381–383.

Pérez de la Vega, M., P. García & R.W. Allard, 1991. Multilocus genetic structure of ancestral Spanish and colonial Californian populations of *Avena barbata*. Proc Natl Acad Sci USA 88: 1201–1206.

Saghai Maroof, M.A., R.M. Biyashev, G.P. Yang, Q. Zhang & R.W. Allard, 1994. Extraordinarily polymorphic microsatellite DNA in barley: Species diversity, chromosomal locations and population dynamics. Proc Natl Acad Sci USA 91: 5466–5470.

Schwartz, D., 1960. Genetic studies of mutant isozymes in maize. Synthesis of hybrid enzymes by heterozygotes. Proc Natl Acad Sci USA 88: 1202–1206.

Turesson, G., 1922. The genotypical response of plant species to the habitat. Hereditas 3: 211–211; 331–348.

Wright, S., 1937. The distribution of gene frequencies in populations. Proc Natl Acad Sci USA 23: 307–320.

Zhang, Q., B.Z. Shen, X.K. Dai, M.H. Mei, M.A. Saghai Maroof & Z.B. Li, 1994. Using bulked extremes and recessive class to map genes for photoperiod-sensitive genic male sterility in rice. Proc Natl Acad Sci USA 91: 8675–8679.