The Influence of Linkage and Inbreeding on Patterns of Nucleotide Sequence Diversity at Duplicate Alcohol Dehydrogenase Loci in Wild Barley (*Hordeum vulgare* ssp. *spontaneum*)

Jing-Zhong Lin,[1] Peter L. Morrell and Michael T. Clegg[2]

Department of Botany and Plant Sciences, University of California, Riverside, California 92521

ABSTRACT

Patterns of nucleotide sequence diversity are analyzed for three duplicate alcohol dehydrogenase loci (*adh1–adh3*) within a species-wide sample of 25 accessions of wild barley (*Hordeum vulgare* ssp. *spontaneum*). The *adh1* and *adh2* loci are tightly linked (recombination fraction <0.01) while the *adh3* locus is inherited independently. Wild barley is predominantly self-fertilizing (∼98%), and as a consequence, effective recombination is restricted by the extreme reduction in heterozygosity. Large reductions in effective recombination, in turn, widen the conditions for linkage to influence nucleotide sequence diversity through the action of selective sweeps or background selection. These considerations would appear to predict (1) homogeneity in patterns of nucleotide sequence diversity, especially between closely linked loci, and (2) extensive linkage disequilibrium relative to random-mating species. In contrast to these expectations, the wild barley data reveal heterogeneity in patterns of nucleotide sequence diversity and levels of linkage disequilibrium that are indistinguishable from those observed at *adh1* in maize, an outbreeding grass species.

SELF-FERTILIZATION is a common form of reproduction among grass species. This mode of reproduction has profound genetic consequences, owing to its effect on genotypic frequency distributions. Mendel (1865) first noted that recurrent self-fertilization is expected to reduce the frequency of heterozygotes by one-half per generation. A reduction in the frequency of heterozygotes reduces the effective rate of recombination that in turn leads to wider conditions for the existence of transient or stable linkage disequilibrium (LD). To quantify this effect, Weir and Cockerham (1973) presented an expression for the rate of decay of LD for linked pairs of genes in mixed-mating systems where a fraction *s* of zygotes arise from self-fertilizations and *t* (=1 − *s*) arise from random outcrossing. According to calculations based on this expression, self-fertilization retards the decay of LD for unlinked genes by a factor of ∼1/*t*, where *t* is the outcrossing rate. Thus, for example, a species with 98% self-fertilization would exhibit a 50-fold reduction in the rate of decay of LD between unlinked genes. On the basis of this fact and extensive empirical research, Allard (1975) argued that the interaction between self-fertilization and selection should favor the development of coadapted gene complexes. Furthermore, conditions for the maintenance of genetic diversity by natural selection are much more restrictive in self-fertilizing species (*e.g.*, Stebbins 1950; Hayman 1953; Kimura and Ohta 1973). Even selectively neutral diversity is expected to be restricted owing to background selection and the associated reduction in effective population size in inbreeding species (Charlesworth *et al.* 1993; Nordborg 2000).

Consistent with these theoretical expectations, isozyme surveys of predominantly self-fertilizing plant populations have tended to show (1) reduced levels of polymorphism compared to outcrossing species (Hamrick and Godt 1990) and (2) higher levels of linkage disequilibrium among both linked and unlinked loci (Hastings 1990). It is now possible to examine patterns of LD in species-wide samples of gene sequences where physical linkage is measured in terms of a few hundred or a few thousand nucleotides. To date, most of these studies have concentrated on a few random mating species (Drosophila, humans, and *Zea mays*). The most extensive data derive from studies of human genetic diversity (Reich *et al.* 2001; Stephens *et al.* 2001). Determining the physical distance encompassed by a LD domain in humans is an area of active research and recent results are leading to a revision of earlier suggestions that LD is restricted to at most a few thousand base pairs. For example, Reich *et al.* (2001) report that the half-length for decay of LD is ∼60 kb and in many cases exceeds 80 kb in major human populations. Numerous sampling issues are associated with this and similar studies (Weiss and Clark 2002). Moreover, there

is substantial heterogeneity in estimated LD domains in human populations owing to population structure and historical demographic events (Ardlie *et al.* 2002).

Nevertheless, the recent measurement of LD domains is important because it establishes that substantial LD domains can exist in random-mating species. In contrast, a recent species-wide survey of 21 loci on chromosome 1 of maize estimated that LD decays to the neighborhood of zero for sites separated by 500 bp (Tenaillon *et al.* 2001). This suggests that LD may be extremely limited in species-wide samples from random-mating plant species.

Recent studies of the predominantly self-fertilizing plant *Arabidopsis thaliana* suggest that LD may extend over physical distances >150 kb (Nordborg *et al.* 2002). Estimates of intragenic recombination from sequence data also provide an important indirect measure of the relative importance of recombination in self-fertilizing species. These estimates provide occasional evidence for intragenic recombination in species-wide samples from *A. thaliana* (*e.g.*, Kawabe *et al.* 1997; Kawabe and Miyashita 1999; Kuittinen and Aguadé 2000) and wild barley (*Hordeum vulgare* ssp. *spontaneum*; Lin *et al.* 2001).

In this article, we use a coalescent framework to analyze diversity patterns for three duplicate alcohol dehydrogenase loci (*adh1*, *adh2*, and *adh3*) in the genome of wild barley. Alcohol dehydrogenase (*adh*; EC 1.1.1.1) is an important enzyme in anaerobic metabolism and is usually encoded by a small multigene family in higher plants. All grass genomes so far investigated possess *adh1* and *adh2* loci. Phylogenetic analyses indicate that these two loci duplicated shortly before the origin of the grass family, ~70 million years ago (Gaut *et al.* 1999). A third locus, *adh3*, duplicated from *adh2* ~16 million years ago within the lineage that includes wild barley (Trick *et al.* 1988). The more ancient *adh1* and *adh2* genes are closely linked on chromosome 4 of barley. No recombinants have been observed between *adh1* and *adh2* in experimental crosses leading to an estimate of $r = 0.0$ with a 95% confidence level of $0.0 < r < 0.01$ (Brown 1980; Hart *et al.* 1980). The more recently derived *adh3* locus is freely recombining with *adh1* and *adh2* (Harberd and Edwards 1983; Trick *et al.* 1988; Brown *et al.* 1989).

Wild barley is a diploid ($n = 7$) grass that is distributed throughout the Near East and Far East (ranging from Afghanistan and Turkmenistan through Iran and Turkey into Iraq, Israel, Syria, and Jordan; von Bothmer *et al.* 1995). Wild barley is the progenitor of domesticated barley and as such has been of long-standing interest to students of crop evolution. Isozyme studies of wild barley conducted in the 1970s indicate moderate levels of genetic diversity (*e.g.*, Brown *et al.* 1978a). Marker loci have been used to estimate outcrossing rates from a number of local populations; these average ~1.6% outcrossing or >98% self-fertilization (Brown *et al.* 1978b).

In this article we analyze patterns of DNA sequence diversity and LD between the three functionally similar *adh* loci on the basis of a species-wide sample of 25 accessions from the natural range of wild barley. The data show heterogeneous patterns of diversity between loci. Patterns of LD for *adh1* and *adh2* within loci are essentially identical to those calculated from maize, an outcrossing species. Some significant LD is detected between the tightly linked *adh1* and *adh2* loci. LD is larger in magnitude, but still modest within loci in samples that span the geographic range of this predominantly self-fertilizing species. The modest levels of LD suggest that the randomizing forces of mutation and recombination dominate inbreeding in the determination of LD patterns at the timescales represented by this sample.

MATERIALS AND METHODS

Plant materials: Twenty-five wild barley accessions were used in this study. These accessions, which were also studied for the *ahd1* and *adh3* genes (Cummings and Clegg 1998; Lin *et al.* 2001), represent a species-wide sample covering the natural geographic range of wild barley with no population subsampling. Detailed sources and culture conditions for the plant samples and genomic DNA isolation procedure were described in Lin *et al.* (2001).

Gene nomenclature: Some confusion exists regarding the identification of *adh2* and *adh3*. Trick *et al.* (1988) originally cloned all three *adh* genes from a genomic library of barley and established on the basis of sequence similarity that one of these genes is homologous to maize *adh1*, thus confirming this gene as barley *adh1*. The *adh3* duplication arose subsequent to the separation of the maize and barley lineages so a sequence similarity argument could not be used to discriminate between *adh2* and *adh3*. On the basis of a comparison of null isozyme mutations to the *adh2* and *adh3* sequences, Lin *et al.* (2001) suggested that the identity of *adh2* and *adh3* had been reversed by Trick *et al.* (1988). In this article we adhere to the nomenclature established in Lin *et al.* (2001).

PCR and sequencing: PCR primers were designed on the basis of the published barley *adh3* sequence of Trick *et al.* (1988), which we now conclude to actually correspond to the *adh2* isozyme locus. The gene was amplified as two segments, *1* and *2*, which overlap in exon 4. The total amplified region begins in exon 1 and extends into exon 9 (Figure 1). Templates for DNA sequencing were generated using a two-step nested primer amplification procedure. Segment 1 was first amplified with primers *f1b* (5'-CAAAATCTACGTAGCAAC GAAC-3') and *r1a* (5'-GAGAGACCACAGCTGAGGACA-3'). The resulting PCR products were then used as templates to reamplify this region using two nested primers, *f1d* (5'-ATCC GTCTTTCTCTTCCAGC-3') and *r1b* (5'-GGGGTTGATCTTT GCGAG-3'). Segment 2 was initially amplified with primers *f2a* (5'-ACCGGCGAGTGCAAGGAC-3') and *r2b* (5'-CATGTA CATCTCGACCACTTC-3') and then reamplified with primers *f2b* (5'-ATCGTGGCGTGATGATC-3') and *r2a* (5'-GAAGAA GGTGCCTTTCAGG-3'). PCR amplification was performed as described in Lin *et al.* (2001). The PCR products were cleaned using a PCR purification kit (QIAGEN, Chatsworth, CA). Nucleotide sequences were determined on both strands by direct sequencing of PCR products on an ABI377 automated sequencer (Applied Biosystems, Foster City, CA) by an outside provider (University of Maine).
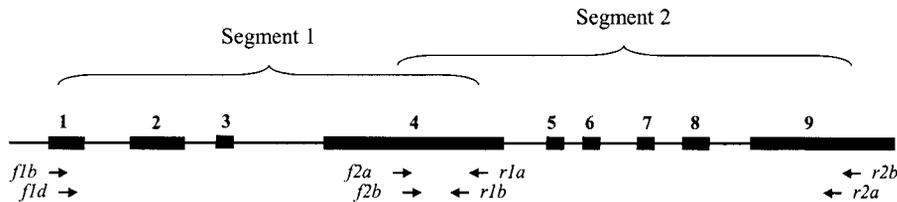
FIGURE 1.—Gene structure of *adh2* in barley. The nine exons are numbered. Segments 1 and 2 were amplified with the primers indicated by arrows.

Sequence analysis: Owing to the high level of self-fertilization in wild barley all accessions studied were homozygous for the *adh2* locus. As a consequence the sampling of accessions is tantamount to sampling haplotypes directly. The sequences of both strands of each accession were aligned to establish a consensus sequence and an alignment for the 25 accessions was built using SEQUENCHER (Gene Codes, Ann Arbor, MI). Estimates of linkage disequilibria and their associated levels of significance were obtained using the program DnaSP (ROZAS and ROZAS 1999), and polymorphism analyses were carried out by using the program SITES (HEY and WAKELEY 1997). Tests of neutrality and determination of their associated significance levels used the programs of FU (1997). Haplotype trees were constructed using statistical parsimony (TEMPLETON *et al.* 1992) as implemented by TCS version 1.13 (CLEMENT *et al.* 2000). Sequence insertions and deletions were treated as a fifth character state. Sequences of *adh1* for the same 25 accessions (CUMMINGS and CLEGG 1998) were retrieved from GenBank and reanalyzed using the methods applied to *adh2* and *adh3*.

RESULTS

Nucleotide sequence polymorphism: The sequenced *adh2* region was 1980 bp, including 946 bp of coding and 1034 bp of noncoding sequence. Nineteen haplotypes are observed in the *adh2* sample (Figure 2). Of the 46 polymorphic sites in the sample, 13 are synonymous, 11 are replacements, 12 are in introns, and 10 represent insertion/deletion events (Table 1). There are substantially larger numbers of singleton polymorphisms in all regions of *adh2* than expected under a neutral model (Table 1).

Figure 2 shows the distribution and types of polymorphic sites in *adh2* for the 25 accessions. There are several major haplotypes, allowing for one or two nucleotide deviations. Haplotype groups I, III, IV, and V account for more than one-half the accessions. Haplotype II and group IV appear to be recombinants. Genealogical analysis showed no evidence for geographic structuring in the *adh2* sample, where the common haplotypes are found across the entire distribution of wild barley (Figure 3).

A 1-bp insertion at position 1240 for accession 24 resulted in early termination in the amino acid translation. We did not observe a null allele at the isozyme level. But with limited isozyme data at hand (not shown), the effect of this insertion is yet to be determined. There is a 365-bp insertion in intron 3 for accession 39 (not shown). The origin of this DNA fragment is unclear. It has ∼42% similarity with a rice genomic DNA clone, but no nucleotide sequence similarity with any of the

alcohol dehydrogenase genes in grasses reported thus far. Thus it could be an interlocus recombinant or a transposable element lacking terminal repeat sequences.

Estimates of nucleotide diversity: Table 2 gives WATTERSON's (1975) θ estimate and TAJIMA's (1989) π estimate of nucleotide diversity and test statistics for deviation from a neutral genealogy for all three genes partitioned by intron, synonymous, and replacement sites. The diversity statistics are clearly heterogeneous with a 10-fold range from *adh1* to *adh3*. *Adh2* is intermediate with about twice the diversity observed at *adh1*. Moreover, the three loci show different test statistic patterns. *Adh1* has significantly negative values for replacement sites and largely nonsignificant negative values for other sites. *Adh2* has negative test statistics and a few are marginally significant. *Adh3* has positive or nonsignificant negative values for most tests except for Tajima's *T* for introns. The positive values at *adh3* are a result of the regional differentiation between two major sequence types observed at this locus (see DISCUSSION).

Estimates of linkage disequilibrium and recombination: Because *adh1* and *adh2* are tightly linked it is of interest to examine the data for intra- and interlocus LD. A standard χ^2 test (with 1 d.f.) for significant pairwise association can be calculated as $NR^2 = ND^2/[p_a(1 - p_a)q_b(1 - q_b)]$, where N is the sample size, D is the standard measure of linkage disequilibrium for a pair of diallelic loci, and p_a and q_b are the marginal site frequencies. The test statistic was calculated for all sites where the minority type was represented two or more times in the sample (that is, singletons were excluded from the calculations of D). A Bonferroni correction for multiple tests was applied to judge significance of the resulting test statistics. Six and 13 polymorphic sites in *adh1* and *adh2*, respectively, satisfy this criterion, yielding 15 and 78 within-locus tests and 78 between-locus tests. Of these, 3 out of 15 tests or 20% were significant beyond the adjusted 5% level within *adh1* and 26 out of 78 tests (33%) were significant within *adh2*. Only 4 between-locus tests were significant (5%). These analyses indicate moderate but far from complete association within loci and relatively little association between the two closely linked loci. Consistent with the statistical analysis of LD, at least two recombination events within the *adh2* region are detected using HUDSON and KAPLAN's (1985) method, one between nucleotide positions 173 and 713 and another between positions 838 and 1710. The population recombination parameters were estimated to be $\gamma = 0.000463$/bp (HEY and WAKELEY 1997)

```
                                      111111111111111111111111
                       1112334567778888888890012234566667777 8888
Nucleotide Position    3334447811663419912345790794770534780 11220134
                       7890123559388433670876762662064478979 506184690

Polymorphism Type      -----SRRRNNNN-NSSSSSSSSSRSSRN-RNRSN-NNRRRSRNN--

Haplotype      Cons.   -----GAGCCGATGGCTCGTCGCGCCCT-CAGCT-ACGCGCCCCC-
Group    Acce.

         11    ......G.......T....................T.........
         25    ......G.......T....................T.........
         44    ......G.......T....................T.........
         13    ......G.......T....................T.........
  I      16    ......G.......T....................T.........
         36    ......G.......T......A.............T.........
         43    ......G.....-T.....T...............T.........
         30    ......G.......T.........T...T....T...........
         10    ......G.......T...................T.........-C

  II     2     AAGGT.........-T..................T...........

         9     AAGGT..C.......AC..CA......C...............-C
         12    ...............AC..CA......................C
         17    ..........C..AC..CA........................C
  III    3     ........G.....AC..CA.......................C
         22    ........G.....AC..CA.......................C
         32    ........A..C..AC..CA.......................C

         35    ...............AC..CA................T....T.C
  IV     21    ...............AC..CA............A.G..T....T.C
         28    ...............C...A............T.....T.T..T.C

         4     .....AG....G.........T.T......T..........-C
  V      6     ......G....G.........T.T......T..........-C
         27    ......G...TG.........T.T......T...........

         39    ....................T.........T.........
  VI     38    ...............T...C..........T......A.C......
         24    ................T..........T...........GT...
```

FIGURE 2.—Polymorphic site distribution in the *adh2* regions among 25 wild barley accessions. Sites in exons are highlighted. Dots denote consensus sequences. Polymorphism types are as follows: S, synonymous; R, replacement; N, noncoding; —, deletion.

and $4Nc = 0.014799$/bp (HUDSON 1987). As noted above, inspection of the haplotype data also reveals evidence for several recombinant genotypes (see Figure 2). For example, accession 2 appears to be a recombinant between haplotype group I and accession 9 of haplotype group III. Haplotype group IV also appears to be a recombinant between haplotype group III and other haplotypes not detected in this sample. However, it is difficult to determine the boundary of linkage blocks and their associated level of significance due to the large number of haplotypes involved.

The power to detect LD within and between loci is limited owing to the very low levels of polymorphism at *adh1* and the weak statistical power of these tests (BROWN 1975). Thus the detection of any statistically significant LD between loci following Bonferroni correc-

TABLE 1

Number of polymorphic sites observed and expected under a neutral model (in parentheses) at various frequencies in regions of *adh2* in wild barley

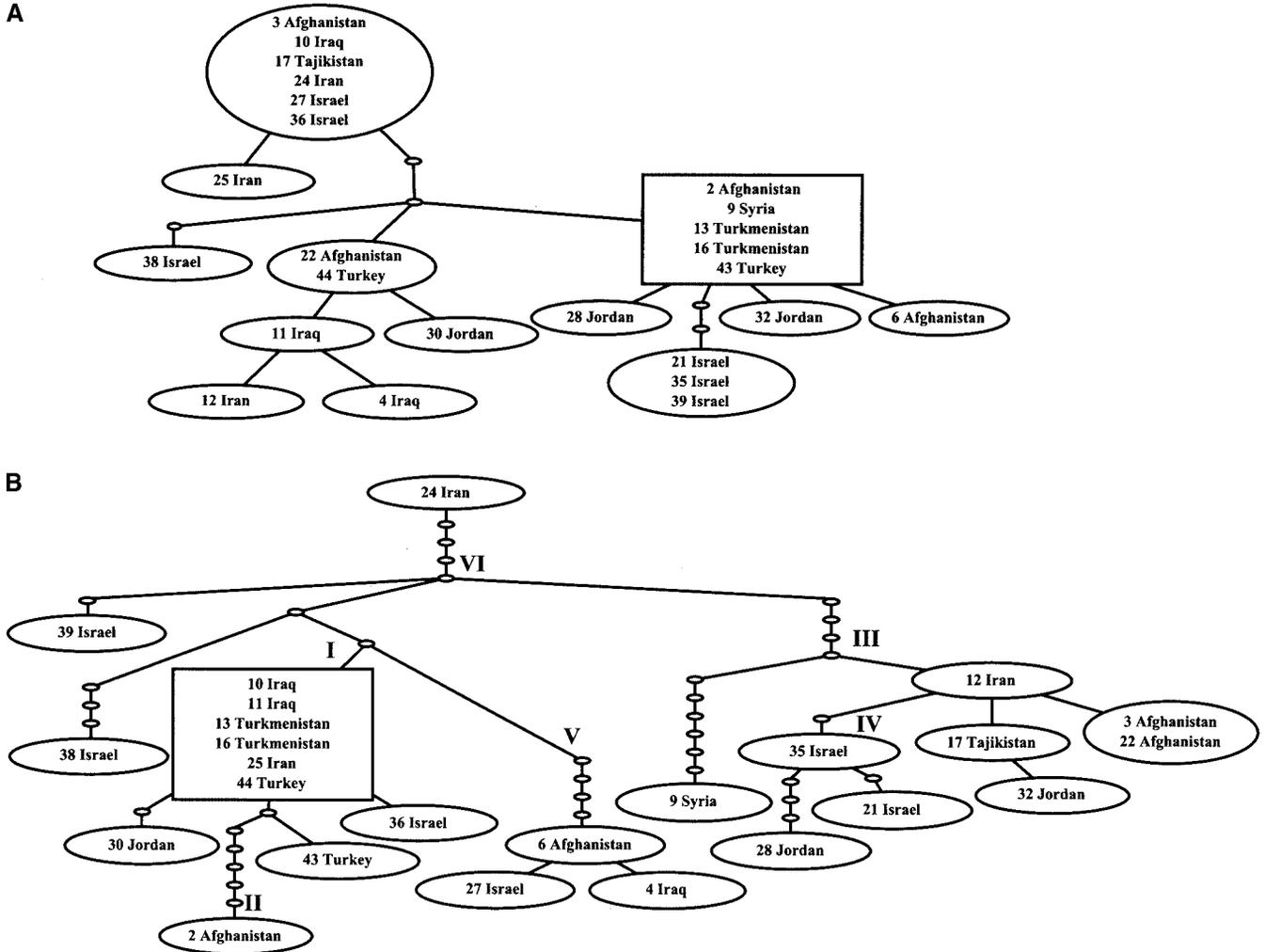| Region	Total	No. of polymorphic sites at frequency of											
		1	2	3	4	5	6	7	8	9	10	11	12
All	36	22 (10.0)	2 (5.2)	6 (3.7)	0 (2.9)	0 (2.4)	0 (2.1)	0 (1.9)	2 (1.8)	2 (1.7)	0 (1.6)	1 (1.6)	1 (1.5)
Intron	12	8 (3.3)	1 (1.7)	2 (1.2)	0 (1.0)	0 (0.8)	0 (0.7)	0 (0.6)	0 (0.6)	0 (0.6)	0 (0.5)	1 (0.5)	0 (0.5)
Exon	24	14 (6.7)	1 (3.5)	4 (2.4)	0 (1.9)	0 (1.6)	0 (1.4)	0 (1.3)	2 (1.2)	2 (1.1)	0 (1.1)	0 (1.0)	1 (1.0)
Replacement	11	7 (3.1)	1 (1.6)	2 (1.1)	0 (0.9)	0 (0.7)	0 (0.6)	0 (0.6)	0 (0.5)	0 (0.5)	0 (0.5)	0 (0.5)	1 (0.5)
Synonymous	13	7 (3.6)	0 (1.9)	2 (1.3)	0 (1.0)	0 (0.9)	0 (0.8)	0 (0.7)	2 (0.6)	2 (0.6)	0 (0.6)	0 (0.6)	0 (0.6)

FIGURE 3.—Statistical parsimony gene genealogy of *adh1* (A) and *adh2* (B) for 25 wild barley accessions as estimated by TCS version 1.13. Accession numbers are followed by the country of origin for each sample. Major *adh2* haplotype groups are denoted by Roman numerals. All branches represent a single mutational step. The haplotypes with the highest outgroup probabilities are displayed as rectangles. Small, unlabeled ovals represent haplotypes inferred from mutational changes, but not sampled in this data set.

tion establishes that significant levels of LD persist within loci and to a much lesser extent between the *adh1* and *adh2* loci. On the other hand, LD between *adh2* and *adh3* (tests not shown) and between *adh1* and *adh3* were not significant (LIN *et al.* 2001), lending further support to our previous finding that the identity of *adh2* and *adh3* had been reversed by TRICK *et al.* (1988).

Figure 4 plots LD as a function of distance (base pairs) for the three barley genes (measured as the squared correlation in allelic state, R^2, defined above) for maize *adh1*. This plot contrasts patterns of R^2 within the barley genes to that observed for an outcrossing grass species. The patterns of LD for barley *adh1* and *adh2* are essentially indistinguishable from maize *adh1*. Barley *adh3* shows very high R^2 across the nearly 2000-bp range of the plot. When the *adh3* data are partitioned into the distinct clusters (see LIN *et al.* 2001), the patterns of LD

become indistinguishable from the other loci (data not shown), establishing that the high LD is accounted for by the extreme population subdivision observed for *adh3*.

DISCUSSION

Patterns of nucleotide sequence diversity are heterogeneous among loci: To contrast the *adh2* data to those of *adh1* and *adh3*, we begin by reviewing several salient features of the *adh1* data set. CUMMINGS and CLEGG (1998) determined *adh1* sequence diversity for a 1362-bp region of the gene in a sample of 45 accessions of wild barley from throughout the range of the species. The total data included 786 bp of exon sequence and 576 bp of intron sequence. A partition of the polymorphic nucleotide sites into synonymous and replacement

TABLE 2

Estimates of nucleotide diversity and test statistics for selection at *adh1*, *adh2*, and *adh3* for 25 accessions of wild barley

Region	Length (bp)	θ/bp	π/bp	T	D^*	F	F^*
			adh2				
All	1971.9	0.0048	0.0032	−1.216	−1.991**	−2.242**	−1.890**
Intron	1034.0	0.0031	0.0017	−1.434**	−1.967**	−2.211**	−1.946**
Exon	936.0	0.0068	0.0049	−0.989	−1.743	−1.916**	−1.630
Replacement	719.5	0.0041	0.0023	−1.337	−1.781	−2.011**	−1.774
Synonymous	216.5	0.0159	0.0133	−0.539	−1.344	−1.374	−1.185
			adh1[a]				
All	1359.9	0.0029	0.0021	−0.926	−1.705	−1.828**	−1.582
Intron	574.2	0.0032	0.0027	−0.436	−0.674	−0.747	−0.648
Exon	786.0	0.0027	0.0016	−1.174	−2.169**	−2.249**	−2.013**
Replacement	606.4	0.0022	0.0007	−1.858***	−2.890***	−3.022***	−2.776***
Synonymous	179.6	0.0044	0.0049	0.247	−0.192	−0.109	−0.071
			adh3[a]				
All	1842.8	0.0149	0.0219	1.734**	0.348	0.931	0.858
Intron	921.9	0.0161	0.0257	2.183***	0.929	1.599	1.437
Exon	929.1	0.0137	0.0179	1.127	−0.350	0.073	0.127
Replacement	715.0	0.0082	0.0088	0.279	−0.746	−0.588	−0.462
Synonymous	212.1	0.0325	0.0488	1.759**	0.038	0.636	0.621

θ, Watterson's estimate; π, Tajima's estimate; T, Tajima's *D*-test; D^*, Fu and Li's *D*-test without outgroup information; F, Fu and Li's *F*-test; F^*, Fu and Li's *F**-test. **$0.01 < P < 0.05$; ***$P < 0.01$.

[a] Data from LIN *et al.* (2001).

sites revealed seven replacement polymorphisms and four synonymous polymorphisms in exons ($\theta = 0.0032$/bp for the entire 45-accession sample). All seven replacement polymorphisms were unique in the sample (each occurred in only a single accession). TAJIMA's (1989) test and analogous tests by FU and LI (1993) and FU (1997) indicated a significant excess of low-frequency amino acid polymorphisms in the sample. Silent polymorphisms in introns and synonymous polymorphisms

in exons were not significant as judged by these tests. The most plausible interpretation of these data is that *adh1* has been subject to a history of strong purifying selection. Low-frequency replacement polymorphisms are likely to be the result of a selection/mutation balance in small local populations where the selection intensity $S < 1/N_e$ (and where N_e is the local population effective size).

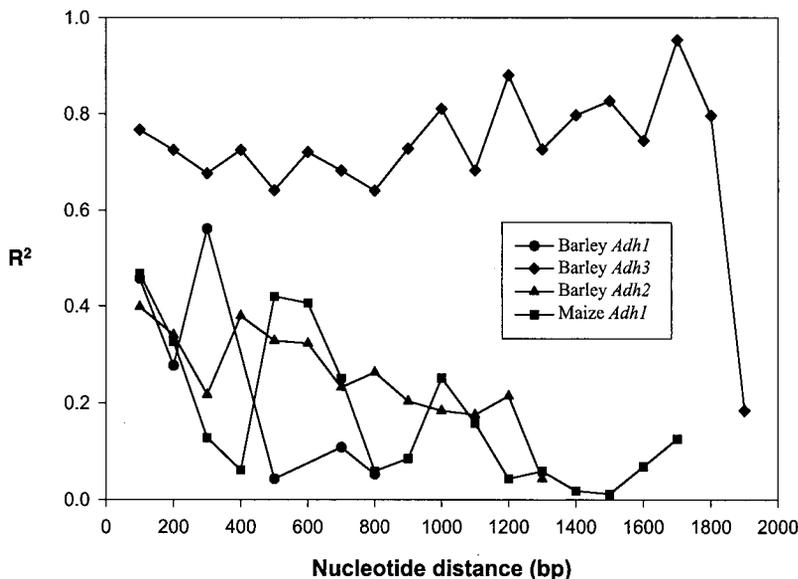The pattern at *adh2* is similar to *adh1* in demonstra-

FIGURE 4.—The relationship between linkage disequilibrium (R^2) and nucleotide distance at *adh1*, *adh2*, and *adh3* in wild barley and *adh1* in maize. The R^2 values are an average of all data points within the distance intervals in 100-bp increments. Intervals with no polymorphic sites are ignored.

ting an overall excess of low-frequency polymorphisms as judged by the test statistics (Table 2). However, the pattern at *adh2* also differs in several important respects. First, the level of nucleotide sequence diversity is 1.6-fold higher at *adh2* ($\theta = 0.0048 \pm 0.0008$) than at *adh1* ($\theta = 0.0029 \pm 0.0008$, for the subset of 25 accessions sampled for *adh2* and *adh3*); second, the excess of low-frequency polymorphisms appears to be associated not only with replacement sites but also with intron and synonymous sites; and third, several replacement polymorphisms occur more than once in the sample (*e.g.*, sites 173, 215, 1404, and 1710). Of particular note is the highly polymorphic site 173 associated with haplotype groups I and V. This isoleucine-to-valine substitution is not associated with any known isozyme variant.

The pattern of replacement polymorphism at *adh2* stands in marked contrast to that observed at *adh1* in that it is not consistent with the simple mutation/selection balance hypothesis invoked to explain the *adh1* pattern. It may be that selective constraints at *adh2* are relaxed relative to *adh1* and these sites are more nearly neutral. This explanation is consistent with the observation that *adh2* shows accelerated rates of protein evolution over the 60- to 70-million-year history of the grass family (GAUT *et al.* 1999). Assuming weaker selective constraints, it is noteworthy that strong purifying selection at *adh1* (background selection) has not reduced θ (the effective population size) at the nearby *adh2* locus to a level comparable to that observed at *adh1*. Put differently, despite a high level of self-fertilization and tight linkage between these two loci combined with evidence for strong purifying selection at *adh1*, the genealogical histories of the two loci are not strongly correlated. An alternative explanation is that the site 173 polymorphism is maintained by selection, but no compelling evidence supports this hypothesis.

LIN *et al.* (2001) analyzed the identical sample of 25 accessions for *adh3* and these data reveal a very different pattern of polymorphism from that observed at *adh1* or *adh2*. For *adh3*, the sample is composed of two dimorphic sequence types in roughly equal frequency that differed from one another at the ~2% level. The estimate of θ is more than fivefold lower at *adh1* than that observed for *adh3* and the test statistics are strongly positive. The estimate of the most recent common ancestor for the *adh3* genealogy is ~3 million years. Several recombinants between the common sequence types were observed in the *adh3* sample and the recombination events were estimated to have occurred at 770,000 and 320,000 years ago. Most importantly, a strong geographic correlation exists with one sequence type occurring almost exclusively in the Far East and the other exclusively in the Near East. The recombinant types are found in samples from the Zagros mountain area, at the point of contact between these two regions (LIN *et al.* 2001). BADR *et al.* (2000) reported a similar geographic pattern for a *Bkn*-3 locus in wild barley and limited

sequence samples from at least one other locus in wild barley also suggest the existence of highly diverged sequence types (P. MORRELL and M. CLEGG, unpublished data). In contrast, no geographic correlation is evident in either the *adh1* or the *adh2* data. The common haplotypes observed at these loci appear to be geographically widespread (see Figure 3). Thus the genealogical history for *adh3* is completely distinct from that observed at the two other unlinked *adh* loci.

Patterns of linkage disequilibrium within and among *adh* loci: We may conclude that each of the three duplicate *adh* loci is subject to the influence of different evolutionary forces, despite the effect of linkage and inbreeding. The detection of intralocus recombination within *adh2* and *adh3* is noteworthy in view of the >50-fold reduction in effective recombination expected in this species. This finding makes it clear that the randomizing effect of recombination operates even when inbreeding is the predominant mode of reproduction. Still LD within loci is moderate; very high LD is observed within the *adh3* sequence sample owing to the two divergent sequence types that characterize the sample (LIN *et al.* 2001). Moreover, moderate LD is detected within *adh1* and *adh2*, corresponding to major haplotypes. However, the levels of LD are essentially indistinguishable from those observed at the homologous *adh1* locus in maize (Figure 4).

What are the possible explanations for the similarity between the maize and barley data? One potential explanation is that wild barley transitioned to self-fertilization from an outcrossing mating system relatively recently in its evolutionary history. There is no compelling evidence one way or the other with which to evaluate this hypothesis, although the closely related species *H. bulbosum* is self-incompatible. A second possible explanation posits biases in the measures of association that obscure real differences. In two commonly used measures of association, R^2 and D', defined as D/D_{max}, D_{max} is the maximum value of D for a given set of gene frequencies. D' has two serious drawbacks; first, it is extremely sensitive to variation in allele frequency, especially when one allele is in low frequency (as is commonly the case in this data set). This means that the variance is very large in small samples. Second, the sampling distribution of D' is unknown [see discussions in WEISS and CLARK (2002) and NORDBORG and TAVARE (2002) for a further critique of D' measures]. The other common measure, R^2, is also a function of nucleotide frequencies and is therefore sensitive to frequency variation. However, R^2 does not tend to inflate the strength of association for rare polymorphic sites as extremely as D'. How might this affect the barley-maize comparison? The estimated values of θ/bp for *adh1* in maize are nearly 5-fold and 10-fold greater than those for barley *adh2* and *adh1*, reflecting both more polymorphic sites and higher frequencies at the average polymorphic site in maize (GAUT and CLEGG 1993; LIN *et al.* 2001). To the extent that poly-

morphic site frequencies are intermediate in maize, we would expect R^2 to be smaller for the extreme case of complete association. Such a bias would tend to inflate the magnitude of the barley LD relative to maize. Thus variation in nucleotide frequencies between the maize and barley samples does not appear to be an explanation for the virtually equivalent magnitudes of within-locus LD in maize and barley.

A third possible explanation concerns the evolutionary timescale spanned by these data. The estimated time to the most recent common ancestor for *adh2* is 460,000 years based on an estimate of the nucleotide substitution rate of 3.5×10^{-9} sites/year (Lin *et al.* 2001). This time interval is clearly sufficient for interlocus recombinants to appear in the sample and, consequently, for a substantial uncoupling of *adh2* evolution relative to *adh1*. This timescale is also sufficient for intralocus recombination to have had a marked randomizing effect based on the observation of recombination between *adh2* haplotypes in the sample. In considering timescale it is important to recall that a species-wide sample was analyzed in this study. The temporal history of a species-wide sample is likely to be much deeper than that for a local population. As a consequence, the expected number of recombination events may be substantial, even over distances of a few hundred nucleotides. (Of course, a past history of higher outcrossing would also enhance the number of recombination events within genes.) In contrast, local populations are ephemeral and typically have much smaller effective population sizes so the expected number of recombination events between closely linked genes drawn from within a local population sample should be greatly reduced. Moreover, because selection is generally dependent on local environmental conditions, it is expected to reduce haplotype diversity leading to increases in LD that may persist for considerable periods of time in self-fertilizing species. Accordingly, we expect hitchhiking effects within genomes of predominantly self-fertilizing species to be substantial at the local population level and to perturb large chromosomal regions. The global averaging associated with species-wide samples largely obscures these local effects.

A recent study of *A. thaliana*, which is believed to have a rate of self-fertilization of ~99%, reported LD domains that appear to extend beyond 150 kb (Nordborg *et al.* 2002). This study was based on a global sample of 20 accessions and was designed to investigate larger chromosomal regions. Several sampling strategies were used, but most pertinent to this discussion was an effort to study a 250-kb region by sequencing 13 short segments scattered throughout the region. There is considerable scatter in the reported data for values of R^2 at very short distances with many data points appearing to fall at or below $R^2 = 0.2$ at the shortest distances. Indeed, the distribution of R^2 appears to be almost uniform over the range from 0 to 100 kb and to have an average <0.2 over this range (see Nordborg *et al.* 2002, Figure 1).

Setting aside the question of multiple tests, $NR^2 > 3.84$ for a critical region of 0.05, which implies $R^2 < 0.192$. Thus it appears that one-half or fewer of the tests would be significant at the shortest distances in the *A. thaliana* data set.

The question emerges whether there is a discrepancy between the results reported here and those of Nordborg *et al.* (2002). Both *A. thaliana* and wild barley have similar nucleotide site diversity levels, so this is unlikely to be a major factor in any differences in levels of LD between these two species. *A. thaliana* is believed to have experienced recent population expansions so demographic influences might account for some differences, although little is known about the demographic history of wild barley. What is more striking in both data sets is their agreement in reporting modest LD at very short distances. We conclude that recombination is a powerful influence even in species where the rate of recombination is substantially reduced through the mating system. In contrast to species-wide patterns, strongly correlated patterns of LD may be expected within local populations where time intervals are short and opportunities for recombination are severely restricted.

LITERATURE CITED

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

Ardlie, K. G., L. Kruglyak and M. Seielstad, 2002 Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet. **3**: 299–309.

Badr, A., K. Müller, R. Schäfer-Pregl, H. El Rabey, S. Effgen *et al.*, 2000 On the origin and domestication history of barley (*Hordeum vulgare*). Mol. Biol. Evol. **17**: 499–510.

Brown, A. H. D., 1975 Sample sizes required to detect linkage disequilibrium between two or three loci. Theor. Popul. Biol. **8**: 184–201.

Brown, A. H. D., 1980 Genetic basis of alcohol dehydrogenase polymorphism in *Hordeum spontaneum*. J. Hered. **70**: 127–128.

Brown, A. H. D., E. Nevo, D. Zohary and O. Dagan, 1978a Genetic variation in natural populations of wild barley (*Hordeum spontaneum*). Genetics **49**: 97–108.

Brown, A. H. D., D. Zohary and E. Nevo, 1978b Outcrossing rates and heterozygosity in natural populations of *Hordeum spontaneum* Koch in Israel. Heredity **41**: 49–62.

Brown, A. H. D., G. J. Lawrence, M. Jenkin, J. Douglass and E. Gregory, 1989 Linkage drag in backcross breeding in barley. J. Hered. **80**: 234–239.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134**: 1289–1303.

Clement, M., D. Posada and K. A. Crandall, 2000 TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9**: 1657–1659.

Cummings, M. P., and M. T. Clegg, 1998 Nucleotide sequence diversity at the alcohol dehydrogenase 1 locus in wild barley (*Hordeum vulgare* ssp. *spontaneum*): an evaluation of the background selection hypothesis. Proc. Natl. Acad. Sci. USA **95**: 5637–5642.

Fu, Y. X., 1997 Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics **147**: 915–925.

Fu, Y. X., and W. H. Li, 1993 Statistical tests of neutrality of mutations. Genetics **133**: 693–709.

GAUT, B. S., and M. T. CLEGG, 1993 Molecular evolution of the *Adh1* locus in the genus *Zea*. Proc. Natl. Acad. Sci. USA **90:** 5095–5099.

GAUT, B. S., A. S. PEEK, B. R. MORTON and M. T. CLEGG, 1999 Patterns of genetic diversification within the Adh gene family in the grasses (Poaceae). Mol. Biol. Evol. **16:** 1086–1097.

HAMRICK, J. L., and M. J. W. GODT, 1990 Allozyme diversity in plant species, pp. 43–63 in *Plant Population Genetics, Breeding, and Genetic Resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. T. KAHLER and B. S. WEIR. Sinauer, Sunderland, MA.

HARBERD, N. P., and K. J. R. EDWARDS, 1983 Further studies on the alcohol dehydrogenases in barley—evidence for a 3rd alcohol-dehydrogenase locus and data on the effect of an alcohol-dehydrogenase-1 null mutation in homozygous and in heterozygous condition. Genet. Res. **41:** 109–116.

HART, G. E., A. K. M. R. ISLAM and K. W. SHEPHERD, 1980 Use of isozymes as chromosome markers in the isolation and characterization of wheat-barley chromosome addition lines. Genet. Res. **36:** 311–325.

HASTINGS, A., 1990 The interaction between selection and linkage in plant populations, pp. 163–180 in *Plant Population Genetics, Breeding, and Genetic Resources*, edited by A. H. D. BROWN, M. T. CLEGG, A. T. KAHLER and B. S. WEIR. Sinauer, Sunderland, MA.

HAYMAN, B. I., 1953 Mixed selfing and random mating when homozygotes are at a disadvantage. Heredity **7:** 185–192.

HEY, J., and J. WAKELEY, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

HUDSON, R. R., 1987 Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50:** 245–250.

HUDSON, R. R., and N. L. KAPLAN, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

KAWABE, A., and N. T. MIYASHITA, 1999 DNA variation in the basic chitinase locus (*ChiB*) region of the wild plant *Arabidopsis thaliana*. Genetics **153:** 1445–1453.

KAWABE, A., H. INNAN, R. TERAUCHI and N. T. MIYASHITA, 1997 Nucleotide polymorphism in the acidic chitinase locus (*ChiA*) region of the wild plant *Arabidopsis thaliana*. Mol. Biol. Evol. **14:** 1303–1315.

KIMURA, M., and T. OHTA, 1973 *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton, NJ.

KUITTINEN, H., and M. AGUADÉ, 2000 Nucleotide variation at the *CHALCONE ISOMERASE* locus in *Arabidopsis thaliana*. Genetics **155:** 863–872.

LIN, J.-Z., A. H. D. BROWN and M. T. CLEGG, 2001 Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). Proc. Natl. Acad. Sci. USA **98:** 531–536.

MENDEL, G., 1865 *Experiments in Plant Hybridisation*, edited by J. H. BENNETT. Oliver & Boyd, London.

NORDBORG, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics **154:** 923–929.

NORDBORG, M., and S. TAVARE, 2002 Linkage disequilibrium: what history has to tell us. Trends Genet. **18:** 83–90.

NORDBORG, M., J. O. BOREVITZ, J. BERGLESON, C. C. BERRY, J. CHORY *et al.*, 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. **30:** 190–193.

REICH, D. E., M. CARGILL, S. BOLK, J. IRELAND, P. C. SABETI *et al.*, 2001 Linkage disequilibrium in the human genome. Nature **411:** 199–204.

ROZAS, J., and R. ROZAS, 1999 DnaSP v. 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

STEBBINS, G. L., 1950 *Variation and Evolution in Plants*. Columbia University Press, New York.

STEPHENS, J. C., J. A. SCHNEIDER, D. A. TANGUAY, J. CHOI, T. ACHARYA *et al.*, 2001 Haplotype variation and linkage disequilibrium in 313 human genes. Science **293:** 489–493.

TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

TEMPLETON, A. R., K. A. CRANDALL and C. F. SING, 1992 A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. Genetics **132:** 619–633.

TENAILLON, M. I., M. C. SAWKINS, A. D. LONG, R. L. GAUT, J. F. DOEBLEY *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA **98:** 9161–9166.

TRICK, M., E. S. DENNIS, K. J. R. EDWARDS and W. J. PEACOCK, 1988 Molecular analysis of the alcohol-dehydrogenase gene family of barley. Plant Mol. Biol. **11:** 147–160.

VON BOTHMER, R., N. JACOBSEN, C. BADEN, R. B. JORGENSEN and I. LINDE-LAURSEN, 1995 *An Ecogeographical Study of the Genus Hordeum*. International Plant Genetic Resources Institute, Rome.

WATTERSON, G. A., 1975 Number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

WEIR, B. S., and C. C. COCKERHAM, 1973 Mixed self and random mating at two loci. Genet. Res. **21:** 247–262.

WEISS, K. M., and A. G. CLARK, 2002 Linkage disequilibrium and the mapping of complex human traits. Trends Genet. **18:** 19–24.

Communicating editor: A. H. D. BROWN