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 $\sim 1/t$, 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., STEGERNS 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; NORDBOG 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 $\sim 60$ 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; Kuitunen 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 ($adh_1$, $adh_2$, and $adh_3$) 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 $adh_1$ and $adh_2$ loci. Phylogenetic analyses indicate that these two loci duplicated shortly before the origin of the grass family, $\sim$70 million years ago (Gaut et al. 1999). A third locus, $adh_3$, duplicated from $adh_2$ $\sim$16 million years ago within the lineage that includes wild barley (Trick et al. 1988). The more ancient $adh_1$ and $adh_2$ genes are closely linked on chromosome 4 of barley. No recombinants have been observed between $adh_1$ and $adh_2$ 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 $adh_3$ locus is freely recombining with $adh_1$ and $adh_2$ (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 Turkmkenistan 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 $\sim$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 $adh_1$ and $adh_2$ within loci are essentially identical to those calculated from maize, an outcrossing species. Some significant LD is detected between the tightly linked $adh_1$ and $adh_2$ 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 $adh_1$ and $adh_3$ 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 $adh_2$ and $adh_3$. 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 $adh_1$, thus confirming this gene as barley $adh_1$. The $adh_3$ duplication arose subsequent to the separation of the maize and barley lineages so a sequence similarity argument could not be used to discriminate between $adh_2$ and $adh_3$. On the basis of a comparison of null isozyme mutations to the $adh_2$ and $adh_3$ sequences, Lin et al. (2001) suggested that the identity of $adh_2$ and $adh_3$ 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 $adh_3$ sequence of Trick et al. (1988), which we now conclude to actually correspond to the $adh_2$ 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'-GAAAATCTAGTACGAAACGAAC-3') and $r1a$ (5'-GAAGAGACACAGCTGAAGCGAC-3'). The resulting PCR products were then used as templates to reamplify this region using two nested primers, $f1d$ (5'-ATCCGCTTCTGTCACGACG-3') and $r1b$ (5'-GAGGGTGTATCTTGCCGAG-3'). The resulting PCR products were then used as templates to reamplify this region using two nested primers, $f2a$ (5'-ACGGCGAGTCAACGACG-3') and $r2a$ (5'-GAAGAAGTTCGCTTCTCAGG-3'). Segment 2 was initially amplified with primers $f2b$ (5'-ATGCCGCTGAGTTGATGACG-3') and $r2b$ (5'-CATGTAAGCAGCTACGACCTT-3') and then reamplified with primers $f2b$ (5'-ATGCCTGACGAGTACGACCT-3') and $r2a$ (5'-GAAGAAGTTCGCTTCTCAGG-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).
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 SEQUENCER (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 χ² test (with 1 d.f.) for significant pairwise association can be calculated as $N D^2 = N D / \left( \frac{p_0 (1 - p_0) q_0 (1 - q_0)}{N} \right)$, where *N* is the sample size, *D* is the standard measure of linkage disequilibrium for a pair of diallelic loci, and *p* and *q* 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 $\gamma = 0.000463 / \text{bp}$ (HEY and WAKELEY 1997).
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 \(4N_e = 0.014799/\text{bp} \) (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

<table>
<thead>
<tr>
<th>Region</th>
<th>Total</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>36</td>
<td>22 (10.0)</td>
<td>2 (5.2)</td>
<td>6 (3.7)</td>
<td>0 (2.9)</td>
<td>0 (2.4)</td>
<td>0 (2.1)</td>
<td>0 (1.9)</td>
<td>2 (1.8)</td>
<td>2 (1.7)</td>
<td>0 (1.6)</td>
<td>1 (1.6)</td>
<td>1 (1.5)</td>
</tr>
<tr>
<td>Intron</td>
<td>12</td>
<td>8 (3.3)</td>
<td>1 (1.7)</td>
<td>2 (1.2)</td>
<td>0 (1.0)</td>
<td>0 (0.8)</td>
<td>0 (0.7)</td>
<td>0 (0.6)</td>
<td>0 (0.6)</td>
<td>0 (0.6)</td>
<td>0 (0.6)</td>
<td>0 (0.5)</td>
<td>0 (0.5)</td>
</tr>
<tr>
<td>Exon</td>
<td>24</td>
<td>14 (6.7)</td>
<td>1 (3.5)</td>
<td>4 (2.4)</td>
<td>0 (1.9)</td>
<td>0 (1.6)</td>
<td>0 (1.4)</td>
<td>0 (1.3)</td>
<td>2 (1.2)</td>
<td>2 (1.1)</td>
<td>0 (1.1)</td>
<td>0 (1.0)</td>
<td>1 (1.0)</td>
</tr>
<tr>
<td>Replacement</td>
<td>11</td>
<td>7 (3.1)</td>
<td>1 (1.6)</td>
<td>2 (1.1)</td>
<td>0 (0.9)</td>
<td>0 (0.7)</td>
<td>0 (0.6)</td>
<td>0 (0.6)</td>
<td>0 (0.5)</td>
<td>0 (0.5)</td>
<td>0 (0.5)</td>
<td>0 (0.5)</td>
<td>1 (0.5)</td>
</tr>
<tr>
<td>Synonymous</td>
<td>13</td>
<td>7 (3.6)</td>
<td>0 (1.9)</td>
<td>2 (1.3)</td>
<td>0 (1.0)</td>
<td>0 (0.9)</td>
<td>0 (0.8)</td>
<td>0 (0.7)</td>
<td>2 (0.6)</td>
<td>2 (0.6)</td>
<td>0 (0.6)</td>
<td>0 (0.6)</td>
<td>0 (0.6)</td>
</tr>
</tbody>
</table>
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. (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

**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

<table>
<thead>
<tr>
<th>Region</th>
<th>Length (bp)</th>
<th>θ/bp</th>
<th>π/bp</th>
<th>T</th>
<th>D*</th>
<th>F</th>
<th>F*</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1971.9</td>
<td>0.0048</td>
<td>0.0032</td>
<td>-1.216</td>
<td>-1.991**</td>
<td>-2.242**</td>
<td>-1.890**</td>
</tr>
<tr>
<td>Intron</td>
<td>1034.0</td>
<td>0.0051</td>
<td>0.0017</td>
<td>-1.434**</td>
<td>-1.967**</td>
<td>-2.211**</td>
<td>-1.946**</td>
</tr>
<tr>
<td>Exon</td>
<td>965.0</td>
<td>0.0068</td>
<td>0.0049</td>
<td>-0.989</td>
<td>-1.743</td>
<td>-1.916**</td>
<td>-1.650</td>
</tr>
<tr>
<td>Replacement</td>
<td>719.5</td>
<td>0.0041</td>
<td>0.0023</td>
<td>-1.337</td>
<td>-1.781</td>
<td>-2.011**</td>
<td>-1.774</td>
</tr>
<tr>
<td>Synonymous</td>
<td>216.5</td>
<td>0.0159</td>
<td>0.0133</td>
<td>-0.539</td>
<td>-1.344</td>
<td>-1.374</td>
<td>-1.185</td>
</tr>
<tr>
<td>All</td>
<td>1359.9</td>
<td>0.0029</td>
<td>0.0021</td>
<td>-0.926</td>
<td>-1.705</td>
<td>-1.828**</td>
<td>-1.582</td>
</tr>
<tr>
<td>Intron</td>
<td>574.2</td>
<td>0.0032</td>
<td>0.0027</td>
<td>-0.436</td>
<td>-0.674</td>
<td>-0.747</td>
<td>-0.648</td>
</tr>
<tr>
<td>Exon</td>
<td>786.0</td>
<td>0.0027</td>
<td>0.0016</td>
<td>-1.174</td>
<td>-2.169**</td>
<td>-2.249**</td>
<td>-2.013**</td>
</tr>
<tr>
<td>Replacement</td>
<td>606.4</td>
<td>0.0022</td>
<td>0.0007</td>
<td>-1.858***</td>
<td>-2.890***</td>
<td>-3.022***</td>
<td>-2.776***</td>
</tr>
<tr>
<td>Synonymous</td>
<td>179.6</td>
<td>0.0044</td>
<td>0.0049</td>
<td>0.247</td>
<td>-0.192</td>
<td>-0.109</td>
<td>-0.071</td>
</tr>
<tr>
<td>All</td>
<td>1842.8</td>
<td>0.0149</td>
<td>0.0219</td>
<td>1.734**</td>
<td>0.348</td>
<td>0.831</td>
<td>0.858</td>
</tr>
<tr>
<td>Intron</td>
<td>921.9</td>
<td>0.0161</td>
<td>0.0257</td>
<td>2.183***</td>
<td>0.929</td>
<td>1.599</td>
<td>1.437</td>
</tr>
<tr>
<td>Exon</td>
<td>929.1</td>
<td>0.0137</td>
<td>0.0179</td>
<td>1.127</td>
<td>-0.350</td>
<td>0.073</td>
<td>0.127</td>
</tr>
<tr>
<td>Replacement</td>
<td>715.0</td>
<td>0.0082</td>
<td>0.0088</td>
<td>0.279</td>
<td>-0.746</td>
<td>-0.588</td>
<td>-0.402</td>
</tr>
<tr>
<td>Synonymous</td>
<td>212.1</td>
<td>0.0325</td>
<td>0.0488</td>
<td>1.759**</td>
<td>0.038</td>
<td>0.636</td>
<td>0.621</td>
</tr>
</tbody>
</table>

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.

* Data from Lin et al. (2001).

Sites revealed seven replacement polymorphisms and four synonymous polymorphisms in exons (θ = 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-
ning 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 $\theta$ (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 $\sim 2\%$ level. The estimate of $\theta$ 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 $\sim 3$ 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 $D/D_{\text{max}}$, $D_{\text{max}}$ 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 $\theta$/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 \( 3.5 \times 10^{-8} \) sites/year (Liu 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 intralocal 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.

We thank Dr. Mary Darbin for technical assistance. We also thank the Alfred P. Sloan Foundation and National Science Foundation grant DEB-0129247 for partial support of this work.

LITERATURE CITED


