

UC Irvine

UC Irvine Previously Published Works

Title

Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp *spontaneum*) despite high rates of self-fertilization

Permalink

<https://escholarship.org/uc/item/2qd3b784>

Journal

Proceedings of the National Academy of Sciences of the United States of America, 102(7)

ISSN

0027-8424

Authors

Morrell, P L
Toleno, D M
Lundy, K E
et al.

Publication Date

2005-02-01

Peer reviewed

Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization

Peter L. Morrell, Donna M. Toleno, Karen E. Lundy, and Michael T. Clegg[†]

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697-2525

Contributed by Michael T. Clegg, December 28, 2004

High levels of inbreeding cause populations to become composed of homozygous, inbred lines. High levels of homozygosity limit the effectiveness of recombination, and therefore, retard the rate of decay of linkage (gametic phase) disequilibrium (LD) among mutations. Inbreeding and recombination interact to shape the expected pattern of LD. The actual extent of nucleotide sequence level LD within inbreeding species has only been studied in *Arabidopsis*, a weedy species whose global range has recently expanded. In the present study, we examine the levels of LD within and between 18 nuclear genes in 25 accessions from across the geographic range of wild barley, a species with a selfing rate of $\approx 98\%$. In addition to examination of intralocus LD, we employ a resampling method to determine whether interlocus LD exceeds expectations. We demonstrate that, for the majority of wild barley loci, intralocus LD decays rapidly, i.e., at a rate similar to that observed in the outcrossing species, *Zea mays* (maize). Excess interlocus LD is observed at 15% of two-locus combinations; almost all interlocus LD involves loci with significant geographic structuring of mutational variation.

nucleotide polymorphism | population structure | Wall's *B* | interlocus linkage disequilibrium | inbreeding

Under recurrent self-fertilization, the level of heterozygosity decays at the rate of one-half per locus, per generation (1). Thus, within a few generations a self-fertilizing population is expected to be entirely composed of a collection of homozygous lines. An important consequence of this mating system is an extreme reduction in the rate of effective recombination. Consequently, the decay of linkage disequilibrium (LD) will be arrested. This expectation has generally been borne out in studies of natural populations; within-population levels of LD for isozyme polymorphisms are generally higher in populations of self-fertilizing plants than in outcrossers (2, 3). Many plant species have a mixed mating system with a mixture of outcrossing and self-fertilization, where occasional outcross events produce new heterozygous lines, that within a few generations, sort out into homozygous lines (4). Under this scenario, LD decays at a rate that is a function not only of recombination distance but also the level of outcrossing (5–7).

It has long been argued that the evolutionary potential of predominantly self-fertilizing species is limited by both reduced genetic diversity and a reduction in potential for effective recombination (e.g., refs. 8 and 9, reviewed in ref. 10). Recombinational potential is important because linkage drag, where selection acts on the net fitness of advantageous and disadvantageous mutations that are in LD with one another, both retards the rate of fixation of advantageous mutations and leads to the fixation of deleterious mutations. Despite the theoretical possibility of linkage drag, many of our most important crops, such as wheat, barley, beans, and tomatoes, are predominantly self-fertilizing species. Therefore, the empirical measurement of the actual extent of LD domains in predominantly self-fertilizing species is an important and largely unresolved issue. Recent studies of *Arabidopsis thaliana* (11) suggest that LD domains may

be of the order of 200 kb, which translates into ≈ 1 map unit (cM), but little is known about the progenitors of major crop plants.

In this article, we examine the magnitude of LD in wild barley (*Hordeum vulgare* ssp. *spontaneum*), the progenitor of cultivated barley, and a species with a rate of self-fertilization of $\approx 98\%$ (12). The sample investigated is comprised of 25 individuals from across the geographic range of wild barley. We characterize LD at the sequence level for 18 loci where haplotypes have been fully resolved. The majority of loci are mapped, so it is possible to compare LD at known physical or genetic map distances. If the mutations at a locus do not have a random geographic distribution, but instead, vary locally among geographic regions, LD can be elevated both within and between loci (13). Thus, we also examine the contribution of geographic structuring of haplotype variation to inter- and intralocus LD. The resulting analyses of wild barley data reveal levels of intralocus LD that are only slightly greater than observed in maize, an outcrossing species.

Materials and Methods

Plant Materials. Individuals sampled were drawn from across the range of wild barley, an annual grass native to southwest Asia. The accession numbers and geographic origins of samples are shown in Table 3, which is published as supporting information on the PNAS web site; see also refs. 14 and 15).

Sampled Loci. Analyzed data includes nucleotide sequences from nine previously reported loci (14–17) and nine loci reported here, referred to as: *5'Pepc*, *Cbf3*, *Dhn1*, *Dhn4*, *Dhn7*, *Faldh*, *ORF1*, *Stk*, and *Vrn1*. The sequence from the enzymatic *Pepc* locus was reported in Morrell *et al.* (17); *5'Pepc* is an additional 2,017 bp adjacent to, and immediately upstream from, that region. C-repeat binding factor 3 (*Cbf3*) is a functional, nonenzymatic locus that shows increased expression when plants are exposed to low temperatures (18). The dehydrin loci, *Dhn1*, *Dhn4*, and *Dhn7*, all show increased expression during dehydration (19). *Faldh* is an enzymatic locus, glutathione-dependent formaldehyde dehydrogenase, also classified as an alcohol dehydrogenase class III. *ORF1* is an ORF, identified as a putative cleavage stimulation factor 1; *Stk* is a putative serine/threonine kinase (20). Vernalization 1 (*Vrn1*) is a MADS-box (transcription factor) gene involved in controlling flowering time in relation to a period at low temperatures (21, 22).

Sequencing and Data Assembly. PCR amplification, sequencing, and fragment assembly follow the methods described by Morrell *et al.* (17); i.e., primarily direct sequencing of PCR products with

Abbreviations: LD, linkage disequilibrium; FET, Fisher's exact test.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AY895831–AY896053).

[†]To whom correspondence should be addressed at: Department of Ecology and Evolutionary Biology, 321 Steinhaus Hall, University of California, Irvine, CA 92697-2525. E-mail: mclegg@uci.edu.

© 2005 by The National Academy of Sciences of the USA

study), the data sets are otherwise directly comparable. Based on 1,000 permutations of the product moment correlation between distance and r^2 , only 7 of the 21 maize loci show a significant correlation at $P < 0.05$. This result may be because of a relatively small number of parsimony-informative sites at loci that do not show a significant correlation. There are always 13 or fewer sites at such loci.

Wall's B for the 21 maize loci varies from 0 to 0.645 with a mean of 0.207. This value is very similar to the mean value of Wall's B of 0.222 from the 18 wild barley loci.

Intralocus LD at 18 of the maize loci is shown in Fig. 3; we have excluded loci that appear to have been subject to strong selection (*D8*, *Tb1*, and *Ts2*) (41). This results in 1,516 pairwise comparisons. Based on the LOWESS curves in Figs. 2 and 3, the initial level of LD is lower in wild barley than in maize. The rate of decay of significant LD is greater in the wild barley data set than in maize. The mean level of significant LD in wild barley shows the greatest decline over the first 300 bp (Fig. 1), whereas the level of significant LD in the maize data set declines most rapidly after 400 bp (Fig. 2). At 1,000 bp, the two data sets show very similar levels of significant LD.

Discussion

We have examined the level of LD within and between 18 loci from a sample of wild barley from across the species range. Wild barley has a rate of self-fertilization of $\approx 98\%$, which results in a very low level of heterozygosity; reducing dramatically the effective rate of recombination relative to random mating. Although the effective rate of recombination should be reduced, 10 of 18 sampled loci show a significant negative correlation of LD with physical distance in base pairs. The loci where a decline of LD with distance is not evident within the locus have either very low levels of polymorphism (10 or fewer segregating sites, i.e., *Adh1*, α -*amy1*, *Pepc*, *5'Pepc*, and *Vrn1*) or geographic subdivision of haplotypes (i.e., *Cbf3*, *Dhn9*, and *G3pdh*).

Interlocus comparisons of LD were performed by using randomization of haplotypes at two-locus pairs to test for an excess of LD in the empirical data sets relative to random configurations of the haplotypes present at each locus. The majority of closely linked loci show an excess of interlocus LD. Loci linked at a distance of ≥ 7 cM show a level of LD of similar to that at unlinked loci.

A number of factors can contribute to excess interlocus LD, including selection, species-wide reductions in effective population size, or geographic structure. Among these factors, geographic structure provides the most plausible explanation for significant LD between unlinked loci (in the absence of rare epistatic interactions). Under selection or reduced effective population size, the decay of LD is a function of recombination rate and distance. However, LD due to geographic structure is independent of linkage. Approximately 15% of two-locus comparisons demonstrate significant interlocus LD. In the majority

of cases, at least one locus in each of the two-locus pairs had been shown (based on the K_{st}^* or S_{mn} tests) to have significant geographic structure among the three principal geographic regions. Clearly, at the genomic level, large numbers of interlocus associations can be expected simply because of nonrandom spatial distribution of haplotypes. Multilocus associations should occur more frequently in predominantly inbreeding species, because associations among haplotypes that are generated by mutation and random genetic drift can persist in inbred and partially isolated subpopulations (or demes) (2, 43, 44)

Why is relatively low LD observed in wild barley despite an expected 40-fold reduction in the effective rate of recombination in a highly inbreeding species? There are at least three possible explanations. The first concerns the time scale spanned by the data; a species-level sample incorporates all of the history of a locus. For the wild barley *Adh2* locus for example, time to most recent common ancestor (T_{MRCA}) was estimated to be 460,000 years based on observed nucleotide substitutions and a mutation rate of 3.5×10^{-9} sites per year (16). In highly inbred species, effective recombination events are associated with outcrossing events (45). Even with only 2% outcrossing, given this time scale, a substantial number of recombination events could accumulate.

A second possible explanation is that the relatively low level of LD observed in wild barley results from a recent transition in mating system from outcrossing to selfing (16). As noted by Lin *et al.* (16) the closely related species *Hordeum bulbosum* is self-incompatible. If the transition to self-fertilization was relatively recent, say within the last 100,000 years, then many recombination events that occurred before the transition may still be evident in the data, and levels of LD may be reduced relative to an equilibrium situation. A third possibility is that increased chiasmata frequencies may elevate recombination rates within self-fertilizing lineages. An increase in chiasmata frequency in inbreeding species relative to outcrossing relatives has been reported (reviewed in refs. 9 and 46).

Even by taking these possible explanations into account, it is remarkable that LD in wild barley is of essentially the same magnitude as observed in maize. Recent work in *Arabidopsis* (47) also suggests relatively restricted LD domains, although still larger in magnitude than observed in wild barley. It is clear that species-wide LD within selfers can be quite limited, perhaps surprisingly limited, given the high levels of LD reported within populations. The question now is whether wild barley is unusual or whether relatively low levels of LD are a common feature of inbreeding species. If wild barley and *Arabidopsis* are typical, then classical arguments about the evolutionary potential of predominantly self-fertilizing species will need to be revisited.

We thank T. J. Close, A. D. Long, S. J. MacDonald, and S. I. Wright for helpful discussion and A. H. D. Brown, B. S. Gaut, S. Hegde, D. B. Neale, N. Takebayashi, and B. S. Weir for comments on an earlier version of the manuscript. This work was supported by National Science Foundation Grant DEB-0129247.

- Mendel, G. (1866) **4**, 3–47.
- Brown, A. H. D. (1979) *Theor. Popul. Biol.* **15**, 1–42.
- Hastings, A. (1990) in *Plant Population Genetics, Breeding and Genetic Resources*, eds. Brown, A. H. D., Clegg, M. T., Kahler, A. L. & Weir, B. S. (Sinauer, Sunderland, MA), p. 449.
- Allard, R. W. (1975) *Genetics* **79**, 115–126.
- Weir, B. S. & Cockerham, C. C. (1973) *Genet. Res.* **21**, 247–262.
- Golding, G. B. & Strobeck, C. (1980) *Genetics* **94**, 777–789.
- Nordborg, M. (2000) *Genetics* **154**, 923–929.
- Stebbins, G. L. (1950) *Variation and Evolution in Plants* (Columbia Univ. Press, New York).
- Grant, V. (1958) *Cold Spring Harbor Symp. Quant. Biol.* **23**, 337–363.
- Takebayashi, N. & Morrell, P. L. (2001) *Am. J. Bot.* **88**, 1143–1150.
- Nordborg, M., Borevitz, J. O., Bergelson, J., Berry, C. C., Chory, J., Hagenblad, J., Kreitman, M., Maloof, J. N., Noyes, T., Oefner, P. J., *et al.* (2002) *Nat. Genet.* **30**, 190–193.
- Brown, A. H. D., Zohary, D. & Nevo, E. (1978) *Heredity* **41**, 49–62.
- Nei, M. & Li, W. H. (1973) *Genetics* **75**, 213–219.
- Cummings, M. P. & Clegg, M. T. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 5637–5642.
- Lin, J.-Z., Brown, A. H. D. & Clegg, M. T. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 531–536.
- Lin, J.-Z., Morrell, P. L. & Clegg, M. T. (2002) *Genetics* **162**, 2007–2015.
- Morrell, P. L., Lundy, K. E. & Clegg, M. T. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 10812–10817.
- Choi, D. W., Rodriguez, E. M. & Close, T. J. (2002) *Plant Physiol.* **129**, 1781–1787.
- Choi, D. W., Zhu, B. & Close, T. J. (1999) *Theor. Appl. Genet.* **98**, 1234–1247.
- Dubcovsky, J., Ramakrishna, W., SanMiguel, P. J., Busso, C. S., Yan, L. L., Shiloff, B. A. & Bennetzen, J. L. (2001) *Plant Physiol.* **125**, 1342–1353.
- Yan, L., Loukoianov, A., Tranquilli, G., Helguera, M., Fahima, T. & Dubcovsky, J. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 6263–6268.

22. Yan, L. L., Loukoianov, A., Blechl, A., Tranquilli, G., Ramakrishna, W., SanMiguel, P., Bennetzen, J. L., Echenique, V. & Dubcovsky, J. (2004) *Science* **303**, 1640–1644.
23. Ewing, B., Hillier, L., Wendl, M. C. & Green, P. (1998) *Genome Res.* **8**, 175–185.
24. Ewing, B. & Green, P. (1998) *Genome Res.* **8**, 186–194.
25. Gordon, D., Abajian, C. & Green, P. (1998) *Genome Res.* **8**, 195–202.
26. Nickerson, D. A., Tobe, V. O. & Taylor, S. L. (1997) *Nucleic Acids Res.* **25**, 2745–2751.
27. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
28. Thornton, K. (2003) *Bioinformatics* **19**, 2325–2327.
29. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7**, 256–276.
30. Tajima, F. (1983) *Genetics* **105**, 437–460.
31. Wall, J. D. (1999) *Genet. Res.* **74**, 65–79.
32. Rozas, J., Sanchez-DelBarrio, J. C., Messeguer, X. & Rozas, R. (2003) *Bioinformatics* **19**, 2496–2497.
33. Hudson, R. R. (1992) *Mol. Biol. Evol.* **9**, 969–969.
34. Hudson, R. R. (2000) *Genetics* **155**, 2011–2014.
35. Cleveland, W. S. (1981) *Am. Stat.* **35**, 54.
36. Cleveland, W. S. (1979) *J. Am. Stat. Assoc.* **74**, 829–836.
37. Brown, A. H. D. (1975) *Theor. Popul. Biol.* **8**, 184–201.
38. Weir, B. S. (1996) *Genetic Data Analysis II, Methods for Discrete Population Genetic Data* (Sinauer, Sunderland, MA).
39. Kleinhofs, A., Kilian, A., Maroof, M. A. S., Biyashev, R. M., Hayes, P., Chen, F. Q., Lapitan, N., Fenwick, A., Blake, T. K., Kanazin, V., *et al.* (1993) *Theor. Appl. Genet.* **86**, 705–712.
40. Choi, D.-W., Koag, M. C. & Close, T. J. (2000) *Theor. Appl. Genet.* **101**, 350–354.
41. Tenaillon, M. I., Sawkins, M. C., Long, A. D., Gaut, R. L., Doebley, J. F. & Gaut, B. S. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 9161–9166.
42. Tenaillon, M. I., Sawkins, M. C., Anderson, L. K., Stack, S. M., Doebley, J. & Gaut, B. S. (2002) *Genetics* **162**, 1401–1413.
43. Brown, A., Nevo, E. & Zohary, D. (1977) *Nature* **268**, 430–431.
44. Charlesworth, D. (2003) *Philos. Trans. R. Soc. London B* **358**, 1051–1070.
45. Nordborg, M. (1999) in *Statistics in Molecular Biology and Genetics: IMS Lecture Notes-Monograph Series*, ed. Seillier-Moisewitsch, F. (Inst. Math. Stat., Hayward, CA), Vol. 33.
46. Charlesworth, D., Charlesworth, B. & Strobeck, C. (1977) *Genetics* **86**, 213–226.
47. Nordborg, M., Hu, T. T., Ishino, Y., Jhaveri, J., Toomajian, C., Zheng, H., Bakker, E., Calabrese, P., Gladstone, J., Goyal, R., *et al.* (2005) *PLoS Biol.*, in press.