Grinding up Wheat: A Massive Loss of Nucleotide Diversity Since Domestication

A. Haudry, A. Cenci,* C. Ravel,† T. Bataillon,*‡ D. Brunel,§ C. Poncet,* I. Hochu,* S. Poirier,* S. Santoni,* S. Glémin,‖ and J. David**

*UMR Diversité et Adaptation des Plantes Cultivées, Montpellier, France; †UMR Amélioration et santé des plantes, INRA Site de Crouëlle, Clermont-Ferrand, France; ‡BiRC-Bioinformatics Research Center, Department of Genetics and Ecology, University of Aarhus, Aarhus, Denmark; §INRA, Centre National de Génotypage, Evry, France; and ‖Institut des Sciences de l'Evolution, Université Montpellier, Montpellier, France

Several demographic and selective events occurred during the domestication of wheat from the allotetraploid wild emmer (*Triticum turgidum* ssp. *dicoccoides*). Cultivated wheat has since been affected by other historical events. We analyzed nucleotide diversity at 21 loci in a sample of 101 individuals representing 4 taxa corresponding to representative steps in the recent evolution of wheat (wild, domesticated, cultivated durum, and bread wheats) to unravel the evolutionary history of cultivated wheats and to quantify its impact on genetic diversity. Sequence relationships are consistent with a single domestication event and identify 2 genetically different groups of bread wheat. The wild group is not highly polymorphic, with only 212 polymorphic sites among the 21,720 bp sequenced, and, during domestication, diversity was further reduced in cultivated forms—by 69% in bread wheat and 84% in durum wheat—with considerable differences between loci, some retaining no polymorphism at all. Coalescent simulations were performed and compared with our data to estimate the intensity of the bottlenecks associated with domestication and subsequent selection. Based on our 21-locus analysis, the average intensity of domestication bottleneck was estimated at about 3—giving a population size for the domesticated form about one third that of wild *dicoccoides*. The most severe bottleneck, with an intensity of about 6, occurred in the evolution of durum wheat. We investigated whether some of the genes departed from the empirical distribution of most loci, suggesting that they might have been selected during domestication or breeding. We detected a departure from the null model of demographic bottleneck for the hypothetical gene *HgA*. However, the atypical pattern of polymorphism at this locus might reveal selection on the linked locus *Gsp1A*, which may affect grain softness—an important trait for end-use quality in wheat.

Introduction

Domestication events provide good examples of dramatic morphological and genetic modifications occurring on a short evolutionary time scale. These changes reflect demographic and selective events during the adaptation of crops to a wide range of environments, sometimes very different from those of their native area. Small initial population sizes and intense human selection for agronomic traits are thought to have decreased the available genetic diversity of most crop plants (Tanksley and McCouch 1997). Thus, domestication can be seen as a population bottleneck in most crop species (Buckler et al. 2001). Molecular marker–based studies of crop domestication have increased our understanding of the current genetic status of crop species (Salamini et al. 2002), making it possible to identify agronomically useful genes in wild relatives and to introduce these genes into the cultivated gene pool (Septiningsih et al. 2003) and to identify genes involved in the domestication process or in subsequent selection events (Wright et al. 2005).

Wheat was among the first crop to be domesticated 12,000 years ago in the Fertile Crescent (Nesbitt and Samuel 1998; Tanno and Willcox 2006). Tetraploid forms of current domesticated wheats are derived from a wild tetraploid progenitor, identified as the wild emmer *Triticum turgidum* ssp. *dicoccoides* (referred to as *dicoccoides*). This species has an allotetraploid genome (AABB) resulting from spontaneous amphiploidization between the diploid wild wheat *Triticum urartu* (AA genome, Dvorak et al. 1993, 1998) and an unidentified diploid *Aegilops* species (BB genome), the closest current relative of which is *Ae. speltoides* (Dvorak and Zhang 1990; Daud and Gustafson 1996; Khlestkina and Salina 2001). Molecular data suggest that *dicoccoides* is a recent allopolyploid, originating between 0.25 and 1.3 MYA (Mori et al. 1995; Huang et al. 2002; Dvorak and Akhunov 2005). There are still *dicoccoides* populations in the Fertile Crescent and these populations have been studied with amplified fragment length polymorphism (AFLP) and microsatellite markers (Ozkan et al. 2002; Sasanuma et al. 2002; Thuillet et al. 2005). A recent study based on AFLP data identified 2 different genetic taxa within *dicoccoides*—a Western race (Israel, Jordan, Lebanon, and Syria) and a Central-Eastern race (Iran, Iraq, and Turkey) (Ozkan et al. 2005)—but the level of genetic differentiation of these 2 races was not estimated. The *dicoccoides* genotypes from the Central-Eastern group are more closely related to cultivated populations than those of the Western group, suggesting that only this group contributed to the germplasm of domesticated wheat (Ozkan et al. 2002; Mori 2003). Tetraploid wheat domestication seems to have occurred at a single location, in south-eastern Turkey (Mori 2003; Ozkan et al. 2005). This area has been identified as a cradle of crop domestication in the Neolithic era and a probable site for the beginnings of western agriculture (Heun et al. 1997; Nesbitt and Samuel 1998; Lev-Yadun et al. 2000; Salamini et al. 2002).

The first domesticated tetraploid wheat emmer (*Triticum turgidum* ssp. *dicoccum*, referred to as *dicoccum*) has a nonbrittle rachis and a uniform flowering time, lacks grain dormancy, and has larger kernels than the wild *dicoccoides*. Emmer was spread with human migration throughout Europe and Asia and was the most important crop in the

Key words: wheat, domestication, nucleotide diversity, bottleneck, coalescence.

E-mail: haudry@supagro.inra.fr.

Mol. Biol. Evol. 24(7):1506–1517. 2007
doi:10.1093/molbev/msm077
Advance Access publication April 18, 2007

Fertile Crescent until the early Bronze Age, 10,000 BC (Bar-Yosef 1998). Emmer was gradually replaced by a new form of tetraploid wheat (*Triticum turgidum* ssp. *durum*, referred to as *durum*) considered to be the ancestral form of current macaroni wheat. The transition from emmer to modern durum wheat involved the acquisition of free threshing. Major losses of neutral genetic diversity occurred at successive stages in the history of *Triticum turgidum* ssp. In diversity surveys based on microsatellite loci calibrated for their mutation rate, the wild *dicoccoides* was found to have an average effective population size (N_e) of 32,500 (Thuillet et al. 2005). This size corresponds to the effective number of breeders in an ideal Wright–Fisher population. The estimated effective population size of emmer (*dicoccum*) is only half (N_e = 12,000) and that for cultivated *durum* is only a fifth of this value (N_e = 6,000 in old landraces and 1,300 in the most recently improved varieties). These marked decreases in N_e during wheat improvement history illustrate the intensity of the successive bottlenecks in tetraploid wheat evolution.

Zohary (1999) investigated the number of times that the wild progenitors of Neolithic agriculture were domesticated in the Near East. Based on polymorphism and taxonomic information, he concluded that emmer wheat was domesticated only once, consistent with the monophyletic origin of emmer. This domestication event may have continued over a millennium, during which time wild wheat persisted in cultivated fields (Tanno and Willcox 2006). However, Ozkan et al. (2005) recently argued that the origins of domesticated tetraploid wheat are consistent with a scenario involving 2 major lineages still found in *durum* and *dicoccum*.

The history of tetraploid wheat domestication is well documented, but that of common wheat remains incomplete. Bread wheat (*Triticum aestivum* referred to hereafter as *aestivum*), the most widely cultivated wheat today, is a hexaploid form of free-threshing wheat (genome AABBDD). It is thought to have resulted from recent hybridization (no more than 8,000 years ago, according to Nesbitt and Samuel [1996]) between an allotetraploid wheat (AABB) and the diploid (DD) *Aegilops tauschii* var. *strangulata* (Kihara 1944; McFadden and Sears 1946; Dvorak et al. 1998). The sources of the tetraploid AB genomes of *aestivum* remain a matter of debate. If *T. aestivum* shares its A and B genomes with the *T. turgidum* spp. wheats, its allotetraploid progenitor is currently not identified although it is hypothesized that a domesticated form was involved in this cross because the current distribution range of *Ae. tauschii* does not overlap with the distribution of the wild *dicoccoides* (Nesbitt and Samuel 1996). Zohary and Hopf (2000) suggested that the tetraploid *dicoccum* might be the progenitor of *aestivum*, with a Caspian origin for the hybridization with *Ae. tauschii* generating a hulled hexaploid wheat. This hybridization was then followed by the rapid emergence of free-threshing forms. However, as pointed out by Nesbitt and Samuel (1996), several lines of archaeological evidence, including the lack of remains of hulled hexaploid wheats in this area, are inconsistent with this hypothesis. It is therefore possible that free-threshing in hexaploids was directly inherited from free-threshing tetraploids, consistent with genetic evidence for the rapid emergence of free-threshing tetraploids (Salamini et al. 2002). The A and B genomes of *durum* and *aestivum* not only show extended conservation but also have marked differences (Isidore et al. 2005). However, combined polymorphism analyses of tetraploid and hexaploid wheats have not yet been carried out. No conclusive study has yet identified unambiguously the sources of the A and B genomes of bread wheat among tetraploid potential donors. Based on the D genome polymorphism in bread wheat, amphiploidization with *Ae. tauschii* is thought to have occurred at least twice (Dvorak et al. 1998; Giles and Brown 2006), so there may have been at least 2 different tetraploid progenitors. Subsequent gene flow from tetraploid progenitors to hexaploids, as suggested by Caldwell et al. (2004), may have boosted genetic diversity within bread wheat and blurred the genetic evidence for the origin of *aestivum*.

Few studies have been carried out on nucleotide diversity in wheat because the presence of 2 or 3 closely related homologous copies in the genome prevents the direct sequencing of polymerase chain reaction (PCR) products. Nucleotide sequence variation is much less prone to homoplasy than microsatellite polymorphism. It provides a powerful mean of unraveling the evolutionary history of crop plants and reconstructing genealogies in populations. Microsatellite analyses, as described by Thuillet et al. (2005), may underestimate the consequences of bottlenecks for nucleotide diversity because the high mutation rate of microsatellites might have allowed some recovery of diversity since domestication (Vigouroux et al. 2002). Most crops were domesticated around 10,000 years ago and therefore cannot be considered to be at the mutation-drift equilibrium. Consequently, studies of domestication require demographic scenarios for reconstructing gene genealogy. Coalescent theory (Hudson 1990) allows efficiently simulating sequence samples under different scenarios. Statistical tests can then be used to identify the scenario most likely to account for the observed polymorphism patterns of the studied samples (Nordborg 2003). DNA sequences and a coalescent framework have been used to investigate population bottlenecks in humans (Wakeley and Hey 1997) and in maize (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005). Comparisons of the loss of genome-wide diversity between wild and cultivated species for large sets of genes can be used to calibrate a plausible scenario for domestication bottleneck. Nonselected genes should have levels of nucleotide diversity consistent with a genome-wide demographic bottleneck, whereas genes selected during or after domestication would be expected to show a locally more severe decrease in nucleotide diversity (Wright et al. 2005). This contrast can be used to test whether the patterns of diversity at a given candidate locus in a crop and its wild progenitor can be explained by a demographic event alone or by selection during domestication (Wright et al. 2005).

The aim of this study was to characterize genetic diversity in domesticated (*dicoccum*) and cultivated wheats (*durum* and *aestivum*) and their wild tetraploid relative (*dicoccoides*), to try to unravel the evolutionary history of cultivated durum and bread wheats, and to quantify the impact of domestication bottlenecks on genetic diversity. We addressed these issues by investigating the nucleotide diversity of 21 genes in a sample of 101 individuals from

Table 1
Sequenced Genes, Their Chromosome Location, Putative Function, and Sequenced Length

Gene Name	Location (chromosome, bin[a])	Putative Gene Function	Sequence Length (bp)
11B	1BS9	Unknown	692
91A	3AL5	Unknown	1,252
AapA	2AL1	Amino acid permease	1,019
AlperA	6AL	Xanthine/uracil/vitamin C permease	1,169
Bp2A	3AL3	ATP biosynthesis	1,433
Bp3B	3BL10	ATP biosynthesis	511
Bp5A	3AL3	ATP biosynthesis	635
ChsA	5AS7	Chalcone synthase	436
GdhA	5AL10	Unknown	1,234
Gsp1A	5AS7	Grain softness protein	939
Gsp1B	5BS	Grain softness protein	473
HgA	5AS7	Hypothetical gene	847
HiplA	5AS7	Hedgehog-interacting protein	615
MdhA	3AL3	Malate dehydrogenase	845
Mdh4B	3BL10	Malate dehydrogenase	1,491
Mp7A	2AS	SNF2 family N-terminal domain	878
MybA	3AL3	Transcription factor	1,427
MybB	3BL10	Transcription factor	3,372
NrpA	5AS7	Nodulin-related protein (?)	963
PsyA	5A	Phytoene synthase (carotenoid biosynthesis)	598
ZdsB	2B	Lycopene synthase (carotenoid biosynthesis)	891

[a] Location on the chromosome bin map build, as determined with a set of wheat aneuploids and deletion stocks (Qi et al. 2004).

4 taxa corresponding to representative stages in recent wheat evolution (wild, domesticated, cultivated durum, and bread wheats). We used these data to assess the genealogical relationships between the 4 taxa, to provide insight into the origin of cultivated wheats. We then compared the genetic diversity of the wild population with that of its cultivated relatives and used coalescent simulations to quantify bottlenecks associated with wheat domestication and subsequent selection. Finally, we tested whether some of the genes in our sample were selected during domestication.

Materials and Methods

Plant Materials

We used 4 wheat taxa for DNA sequence analysis: the wild *dicoccoides*, the domesticated *dicoccum*, and 2 wheats cultivated today: durum and bread wheat—*durum* and *aestivum*, respectively. For each taxon, we used a core set of individuals representing the highest available levels of allelic diversity. These individuals were chosen to maximize the number of alleles observed at 30 microsatellite loci (David et al. 2003). We sequenced 28 *dicoccoides*, 12 *dicoccum*, 20 *durum*, and 41 *aestivum* individuals. The accession numbers and geographic origins of the samples are shown in supplementary table S1 (Supplementary Material online). The sample sizes of different loci differed because not all loci were successfully amplified or sequenced in all individuals. Finally, we studied the genetic structure of the wild *dicoccoides* population, using a previously reported data set corresponding to the 52 accessions genotyped with 15 microsatellites (Thuillet et al. 2005).

Design of Genome-Specific Primers

The allopolyploid origin of wheat from 2 (*T. turgidum*) or 3 (*T. aestivum*) ancestral genomes prevents direct sequencing. The sequencing of genes in polyploid wheat requires either cloning or the development of genome-specific primers to ensure that only the targeted copy is amplified. We amplified gene fragments with locus- and genome-specific primers, designed as previously described (Ravel et al. 2006), to prevent the amplification of paralogous and homeologous loci. We then tested the genome specificity of amplification systematically on a set of 7 genotypes: 2 AA diploid accessions (*Triticum monococcum* and *T. urartu*), 2 BB-like diploid accessions (*Ae. speltoides* selfed progeny), 2 DD diploid accessions (*Ae. tauschii*), and 1 tetraploid AABB accession (*Triticum turgidum* ssp. *durum* var. *Langdon*). If a single fragment of the expected size was amplified only in individuals with the targeted genome, the complete sample was amplified. Sequences were submitted to GenBank (GenBank accession numbers are listed in supplementary table S2 [Supplementary Material online]).

Loci Sampled

We sequenced 21 gene fragments from 101 individuals: 15 from the A genome—*91A*, *AapA*, *AlperA*, *Bp2A*, *Bp5A*, *ChsA*, *GdhA*, *Gsp1A*, *HgA*, *HiplA*, *MdhA*, *Mp7A*, *MybA*, *NrpA*, and *PsyA*—and 6 from the B genome—*11B*, *BP3B*, *Gsp1B*, *MdhB*, *MybB*, and *ZdsB*. Six of these genes are located in the vicinity of the *hardness locus*: *ChsA*, *HgA*, *HiplA*, *Gsp1A*, *NrpA*, and *Gsp1B* (Chantret et al. 2005). The size of the amplified fragment, the chromosome location and the putative function of the corresponding gene are reported in table 1. The PCR and sequencing conditions used are described in supplementary table S3 (Supplementary Material online).

Statistical Analysis

Sequences were aligned manually with the Staden Package (Staden et al. 2001). Because of recombination events among loci, it is difficult to assess genealogical

relationships among accessions. To get a rough idea of main relationships, we concatenated all loci except *ChsA* (sequence data lacking for *aestivum*) and performed maximum likelihood (ML) reconstruction (model general time reversible + gamma distribution) using the PHYML software (Guindon and Gascuel 2003). Concatenation resulted in many missing data. We therefore used one of the most parsimonious trees (using DNAPARS procedure of the PHYLIP package version 3.6, Felsenstein 2005) as the starting point for ML search, which is more robust to missing data than using distance trees. Five sequences from *T. timopheevii*, an allotetraploid sister species, were used as outgroup.

We used DnaSp version 4.10 (Rozas et al. 2003) to calculate the number of polymorphic sites (S), the number of haplotypes (h), and the nucleotide diversity per site (π) (Tajima 1983) calculated for the whole sequence (π_{total}) and for noncoding and synonymous sites (π_{silent}). Single-locus and multilocus Tajima's D test (Tajima 1989) was performed in each group using J. Hey's HKA software (http://lifesci.rutgers.edu/~heylab/ProgramsandData/Programs/HKA/HKA_Documentation.htm). We investigated the consequences of domestication for diversity in the wheat genome, using the current wild group *dicoccoides* as a proxy for the initial population before domestication. A recent study has suggested that *dicoccoides* may form 2 main populations (Ozkan et al. 2005). We tested for the presence of these 2 main populations 1) by analysis with STRUCTURE software (Pritchard et al. 2000) of the microsatellites data set for the collection of 52 accessions of *dicoccoides* from Thuillet et al. (2005) and 2) by classical Fst statistic analysis between the 2 groups detected by Ozkan et al. (2005) using the Genetix software (Belkhir et al.; http://www.genetix.univ-montp2.fr/genetix/intro.htm). We investigated the distribution of both neutral and nucleotide diversity in *dicoccoides* by carrying out Mantel's correlations of genetic and geographic distances with microsatellite and sequence data using the GenAlEx 6.0 software (Peakall and Smouse 2006).

Demographic Model

We used a simple model of reduction in effective population size (fig. 1), in which a single ancestral population (the wild population) experienced an instantaneous change in effective population size, t generations ago. The bottleneck intensity α was defined as the ratio of the wild population size (N_a) to cultivated population size (N_p). Higher values of α correspond to more severe bottlenecks. We kept the demographic scenario simple by not allowing for an increase in population size after the bottleneck. This approximation has been shown to have little effect on levels of nucleotide diversity as shown by simulations with this type of bottleneck model in maize (Eyre-Walker et al. 1998). Previous studies have shown that diversity after bottleneck scales to the ratio of the size of the bottleneck population (N_b) and the duration of the bottleneck (d), such that the 2 parameters cannot be estimated separately (Eyre-Walker et al. 1998; Tenaillon et al. 2004; Wright et al. 2005). Here, $d = t$, the number of generations since domestication, and the constant parameter is the product $\alpha \times t$. Assuming a shorter duration of the bottleneck will increase α. The choice of this scenario can also be justified due to the relatively short time (on an evolutionary scale) for the recovery of nucleotide polymorphism after domestication and the continuous selection experienced by wheat populations since domestication.

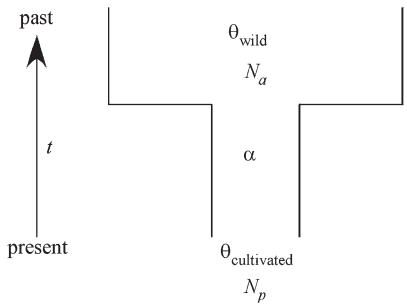

FIG. 1.—Schematic diagram of the coalescent model used in simulations. The ancestral population experienced an instantaneous change in effective population size (N_a), t generations ago. The bottleneck intensity α is defined as the ratio of ancestral population size (N_a) to cultivated population size (N_p).

For the estimation of bottleneck intensities, we used the *dicoccoides* data to calibrate the simulation parameters for the ancestral population, and we used data for *dicoccum* (cultivated emmer), *durum* (durum wheat), or *aestivum* (bread wheat) as the observed data in cultivated groups for a goodness-of-fit analysis (see below). For each locus, the model had 5 parameters (τ, N_a, N_p, θ_{wild}, and $4Nc$):

- τ, the time since the bottleneck was expressed in units of time scaled relative to effective size as $\tau = t/2N_p$, where N_p is the effective population size after the bottleneck. As domestication is thought to have occurred 12,000 years ago (Harlan 1992), t = 12,000. The N_p is equivalent to the ancestral population size (N_a about 30,000; Thuillet et al. 2005) divided by the bottleneck intensity: $N_p = N_a/\alpha$. We therefore used $\tau = \alpha t/2N_a$ in simulations, that is, $\tau = 0.2\alpha$.
- Assuming that *dicoccoides* is the progenitor of the A and B genomes of wheats, we used $\theta_{dicoccoides}$ as a proxy of the initial θ_{wild}. The population mutation rate $\theta_{dicoccoides}$ was estimated by Tajima's π statistic (Tajima 1983), based on sequences from wild *dicoccoides*.
- For each locus, the population recombination parameter ($4Nc$) was estimated from *dicoccoides* data, using Hudson's 2001 method by LDhat program (http://www.stats.ox.ac.uk/~mcvean/LDhat/LDhat1.0/LDhat1.0.html). This parameter was included in simulations when it could be estimated. Otherwise, we assumed that no recombination occurred.

Goodness-of-Fit between Simulations and Observed Data

Coalescent simulations were performed and compared with our data to model the impact of a bottleneck on sequence diversity. Coalescent simulations were run with the "ms" program (Hudson 2002). For each locus and each

cultivated group, 50 values of α were explored on a grid ranging from 1 (i.e., no decrease in effective population size) to 25.5. For each locus and group considered, 10,000 simulations were carried out. Each coalescent simulation was summarized by a π_{simul} and a S_{simul} value. For each scenario, the approximate likelihood of the data at locus i within group j, $L_{ij}(\alpha)$, was calculated as the number of simulations in which both π_{simul} and S_{simul} were within 20% of the observed values of π and S for the data (Weiss and von Haeseler 1998). The intensity of the bottleneck at locus i within group j was estimated as the value α maximizing $L_{ij}(\alpha)$.

We estimated the average bottleneck intensity for each of the 3 cultivated groups by calculating a multilocus likelihood $Lm_j(\alpha)$ as the product over all loci of $L_{ij}(\alpha)$. This approach implicitly assumes that the loci are independent. However, 5 of the loci considered here are located in the same chromosome region, 5AS7 (*ChsA*, *HgA*, *Gsp1A*, *NrpA*, and *HiplA*; Sourdille et al. 1996). We estimated the linkage disequilibrium (LD) between the polymorphic sites for these loci, using TASSEL software (http://www.maizegenetics.net/index.php?page=bioinformatics/tassel/index.html). LD was significant only at the intragene level, and no LD was detected between the different loci of the 5A region ($r^2 < 0.2$ within 1 kb of the "hardness locus," data not shown), so all genes can be assumed to be independent in our likelihood calculation. The intensity of the bottleneck within each group j was calculated as the value of α maximizing $Lm_j(\alpha)$. A 95% confidence interval (CI) was constructed around the estimate of α by identifying the value of α at which the log-likelihood value was 2 log-likelihood units lower than the ML.

Using the Demographic Model to Test for Selection

Selection at some loci would result in the distribution of polymorphisms being skewed at these loci, which might account for the observed variability in bottleneck intensity α among loci. The loci with the most severe bottleneck estimates were considered to be candidate loci for selection during domestication. We investigated whether some loci were outliers in the empirical distribution of most of the loci, by calculating the P value associated with their observed π value. We used the mean value and upper CI limit of α determined by the demographic model to perform additional simulations at each locus. The P value of $\pi_{observed}$ was calculated for these distributions. If significant ($P < 0.05$), the locus was discarded and the analysis was repeated with $n - 1$ loci. This procedure was repeated until no significant effect was detected and for all loci without polymorphism in the cultivated group.

Results

Relationships Between Taxa

Because of low diversity levels, single-locus analyses are not powerful enough to detect clear relationships among the different forms of wheat. We thus performed multilocus analyses combining all genes (fig. 2). Like other tree representation based on a combination of marker information widespread over the genome, this tree should be interpreted with caution because of recombination events between loci. It can give a general picture of accessions relationships but detailed analyses can be misleading. The general topology shows that all cultivated forms are subsets of the wild *dicoccoides* group, consistent with a single domestication event. Three *dicoccoides* accessions fall within cultivated accessions, but we do not have clear explanation for this finding. Long branches are due to the higher diversity in the wild group. The domesticated *dicoccum* forms are dispersed within this cultivated group. Durum wheat individuals cluster together in a subgroup. They are included in the *dicoccum* lineage rather than forming a different lineage. Bread wheat presents a singular pattern, with 2 different groups, 1 lying on an external branch (I) and the other (II) spanning a large proportion of cultivated wheat diversity. To test further this pattern, we performed a STRUCTURE analysis with admixture on the bread wheat data set. We also found 2 clearly distinct groups mainly corresponding to those observed on the tree (see supplementary fig. S4, Supplementary Material online). We found no clear relationship between the genealogy of *aestivum* accessions and their geographic origin (see supplementary table S1, Supplementary Material online).

Genomic Characterization of the Wild *dicoccoides* Group

On the 21,720 bp, corresponding to the 21 loci, in *dicoccoides*, we identified a total of 212 single nucleotide polymorphisms (SNPs). The nucleotide diversity π ranges from 0.0006 (*AapA*) to 0.0116 (*HgA*), with a mean value of 0.0027 (table 2). For both π and θ_W, diversity is greatest for the genes *HgA* ($\pi = 0.0116$ and $\theta_W = 0.0141$) and *ChsA* ($\pi = 0.0113$ and $\theta_W = 0.0112$). The lack of genetic variability made it difficult to estimate the population recombination parameter $4Nc$. We were unable to estimate this parameter for 7 loci (table 2), and it was therefore set to zero when running coalescent simulations (see below). For the other 13 genes, the population recombination rate $4Nc$ ranges from 0.002 (*Mdh4B*) to 0.067 (*ChsA*), with a mean value of 0.015 per nucleotide. When considering each locus individually, we detected no significant departure from the neutral equilibrium model in Tajima's test (Tajima 1989) but values of Tajima's D statistic tend to be overall slightly negative and the multilocus test is highly significant (mean $D = -0.76762$, $P < 0.001$).

Ozkan et al. (2005) found 2 main geographic groups in *dicoccoides* and suggested that domestication occurred likely in the Turkish area. Surprisingly, using the STRUCTURE software (Pritchard et al. 2000), we did not find the same 2 groups. Using sequence data, we found only one group, and using microsatellite data, we only detected a small group of Palestinian and Israeli accessions (but the likelihood of the data assuming 2 populations was only slightly higher than when assuming a single one, supplementary fig. S4 [Supplementary Material online]). In addition, we found no significant correlation between genetic and geographical distances in our samples of *dicoccoides*, either for microsatellite or for sequence data. Significant genetic isolation by distance was detected only between

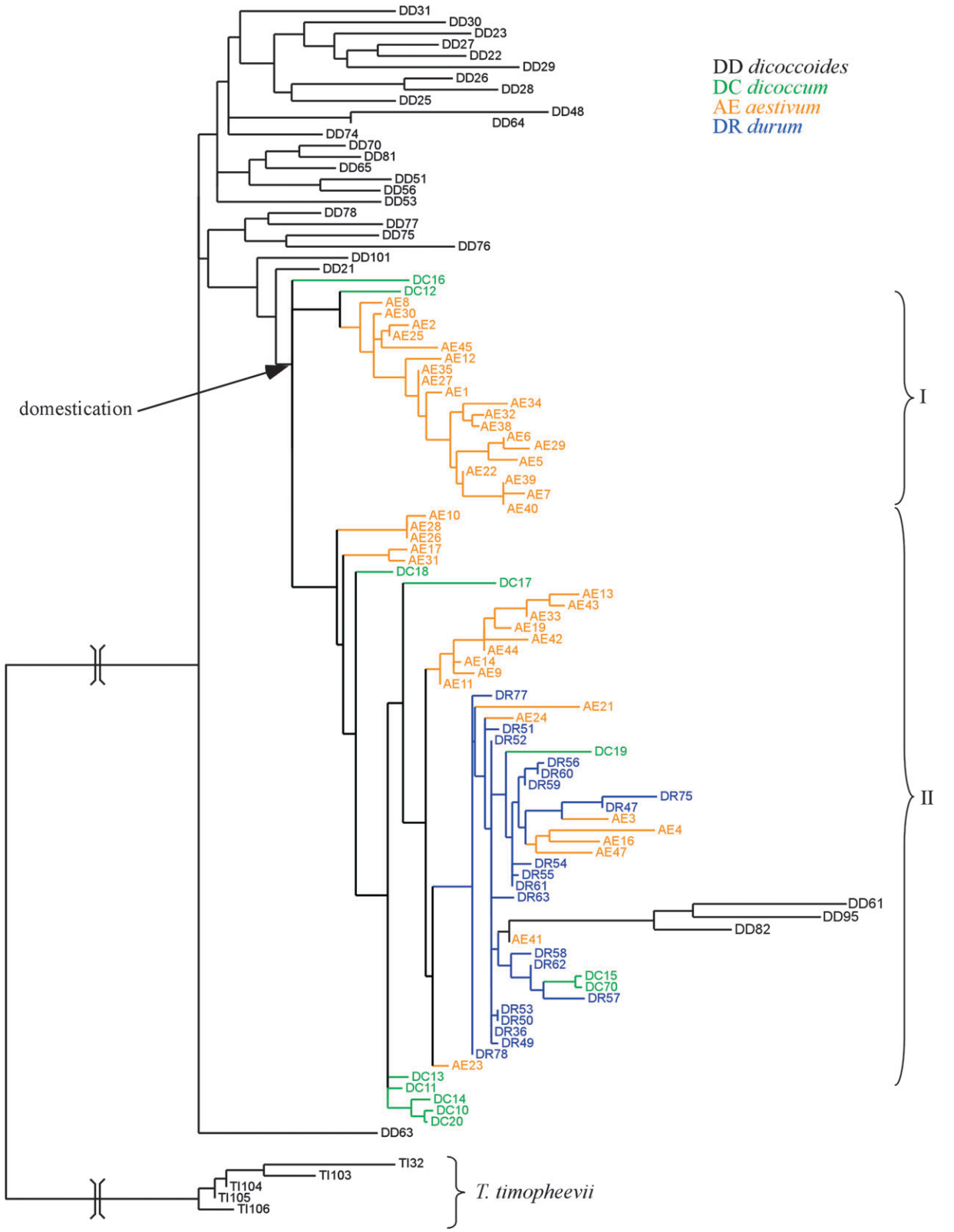

FIG. 2.—Maximum likelihood phylogenetic tree (GTR + Γ) built with sequences for the 20 loci concatenated, with 5 sequences from *T. timopheevii* used as outgroup. Individuals are denoted as shown in supplementary table S1 (Supplementary Material online). The 2 groups of *aestivum* are identified as I and II.

Table 2
Sequence Statistics for the Loci Studied in Wild Emmer

Gene	*n*	*L* (bp)	*S*	$\pi \times 10^{-3}$	$\theta_W \times 10^{-3}$	4*Nc*	Tajima's *D*	π_a/π_s
11B	10	692	10	3.28	5.11	0.0260	−1.59 ns	0.504
91A	10	1,252	7	2.18	1.98	0.0399	0.44 ns	0.061
AapA	9	1,019	2	0.6	0.72	—	−0.58 ns	—
AlperA	10	1,169	7	2.28	2.12	0.0026	0.33 ns	0
Bp2A	27	1,433	18	1.67	3.26	0.0028	−1.72 ns	0.165
Bp3B	10	511	1	1.04	0.69	—	1.3 ns	—
Bp5A	26	635	5	0.83	2.06	—	−1.7 ns	0
ChsA	6	436	11	11.27	11.15	0.0673	0.06 ns	0.39
GdhA	26	1,234	11	2.4	2.45	0.0040	−0.08 ns	0
Gsp1A	25	939	26	4.2	7.33	0.0043	−1.58 ns	0.12
Gsp1B	28	473	7	2.23	3.83	—	−1.25 ns	0.49
HgA	14	847	38	11.63	14.12	0.0047	−1.03 ns	0
HiplA	26	615	6	1.8	2.56	0.0049	−0.87 ns	0
Mdh4B	8	1,491	6	1.72	1.55	0.0020	−0.22 ns	—
MdhA	27	845	7	1.16	2.15	—	−1.4 ns	—
Mp7A	24	878	4	0.97	1.27	—	−0.65 ns	—
MybA	10	1,427	3	0.61	0.74	0.0216	−0.66 ns	1
MybB	10	3,372	16	1.37	1.68	0.0203	−0.86 ns	0
NrpA	17	963	15	2.58	4.61	0.0052	−1.69 ns	0.69
PsyA	28	598	7	1.84	3.02	0.0042	−1.18 ns	0.32
ZdsB	28	891	5	0.82	1.44	—	−1.19 ns	—

NOTE.—*n*, number of *dicoccoides* individuals sampled; *L* (bp), length of aligned sequence, excluding gaps; *S*, total number of segregating sites; $\pi \times 10^{-3}$, average number of pairwise differences calculated on all sites; $\theta_W \times 10^{-3}$, per-site estimates of diversity by Watterson's theta; 4*Nc*, population recombination parameter 4*Nc* (no recombination could be detected for 7 loci [—]); Tajima's *D*, Tajima's *D* statistic for all sites; π_a/π_s, ratio of nonsynonymous diversity (π_a) to synonymous diversity (π_s) (when π_s was zero, it was not possible to calculate the π_a/π_s ratio [—]); ns, not significant.

populations located less than 50 km apart (supplementary fig. S4, Supplementary Material online). However, using the microsatellite data set, we found low but significant *F*st values assigning our accessions to the 2 geographic groups identified by Ozkan et al. (2005) (*F*st = 0.026, *P* value = 0.012 after 1,000 random permutations). Nucleotide diversity is similar in both populations ($\pi = 0.0024$ for the Central East group and $\pi = 0.0027$ for the Western one) and close to the mean diversity obtained over the whole populations ($\pi = 0.0027$). Overall, these results suggest that population structure at a large geographic scale does occur in *dicoccoides* but such a structure is weak and can hardly be detected with 21 genes or 15 microsatellites markers.

From Wild to Cultivated Wheats: Important Losses of Nucleotide Diversity

The transition from wild to cultivated forms was marked by a large decrease in nucleotide diversity. We detected a mean of 10.1 polymorphic sites per locus in the wild *dicoccoides*, whereas only 3 polymorphic sites per locus were found in *dicoccum* and *aestivum* and 1.7 such sites were found in *durum* (supplementary table S5, Supplementary Material online). Nucleotide diversity (π) in the cultivated groups ranged from 0 (3 of 21 genes in *dicoccum*, 7 in *durum*, and 6 in *aestivum*) to 0.003 (*Gsp1B* in *dicoccum*). The mean value of π was 0.0008 for *dicoccum* and *aestivum*, and this value was halved in *durum*. The rate of nucleotide diversity loss was similar when comparing silent sites (measured on noncoding and synonymous sites) and all sites (fig. 3). The domesticated *dicoccum* has 70% less diversity than the wild *dicoccoides*, whereas *durum* is 84% less diverse than the wild taxon. In *aestivum*, nucleotide diversity is 69% lower than that in wild *dicoccoides*. Tajima's *D* statistic tended to be negative in *dicoccum* and *durum* (supplementary table S5, Supplementary Material online). Higher values of Tajima's *D* statistic were obtained in the *aestivum* group (*Gsp1A* and *MybB*).

Coalescent theory–based analyses of nucleotide polymorphism were consistent with strong bottlenecks in each cultivated group (fig. 4). The ML estimate of bottleneck intensity α for the domestication bottleneck (transition *dicoccoides* to *dicoccum*) was $\alpha_{dicoccum} = 3.15$ (CI = 2.07–4.53). The ML estimate of α for the transition from *dicoccoides* to durum wheat was 5.83 (CI = 4.35–7.94), probably due to

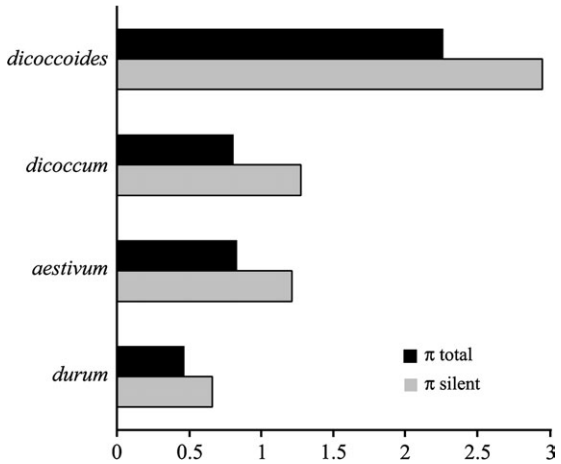

FIG. 3.—Nucleotide diversity (π) expressed in 10^{-3} in the wild (*dicoccoides*) and cultivated groups (*dicoccum*, *durum*, and *aestivum*). Total nucleotide diversity π_{total} was estimated for the whole sequence, whereas π_{silent} corresponds to nucleotide diversity for noncoding and synonymous sites only.

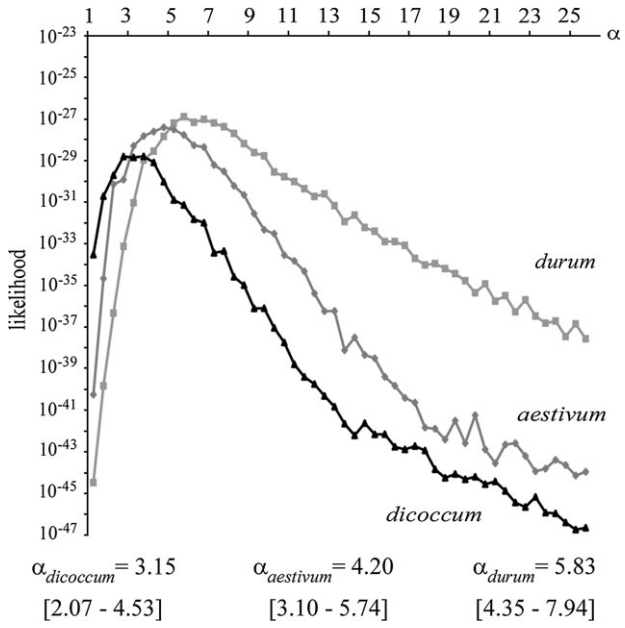

FIG. 4.—Likelihood (*y*axis, logarithmic scale) profiles as a function of bottleneck intensity, α (*x*axis) for each cultivated group, based on the 21 loci ($\alpha = Ne_{dicoccoides}/Ne_{cultivated}$).

a further loss of diversity after domestication. Bread wheat (*aestivum*) displayed a lower loss of diversity after domestication ($\alpha_{aestivum}$ = 4.20, CI = 3.10–5.74). The CI of the bottleneck intensities (α) experienced by the *aestivum* and *dicoccum* groups overlapped considerably. These bottleneck intensities reflect the diversity reduction from wild to domesticated populations when *dicoccoides* is taken as a whole. If only Central-Eastern *dicoccoides* populations were the founders of the domesticated group, true α values for the domestication process per se might be smaller. Therefore, we redid coalescent simulations using the parameters of the Central-Eastern group to characterize the wild initial population. Because, nucleotide diversities are similar to those computed on the whole set of accessions, bottleneck intensities are almost unchanged ($\alpha_{dicoccum}$ = 2.61, α_{durum} = 5.45, and $\alpha_{aestivum}$ = 4.53) and CI overlap (see supplementary fig. S4 [Supplementary Material online] for details).

Variation in Bottleneck Intensity Among Loci

The loss of nucleotide diversity π in cultivated groups compared with the wild *dicoccoides* varied widely among loci (fig. 5). Some genes displayed little or no loss of diversity in the *dicoccum* group (*Gsp1B*, *MybB*, *NrpA*, and *ZdsB*) and in bread wheat (*HiplA* and *MybB*). However, most genes showed a sharp decrease of genetic variability compared with the wild *dicoccoides*. In particular, the *HgA* locus (circled points in fig. 5) displayed a drastic loss of diversity from the wild to the 3 cultivated groups. Some genes were monomorphic in the cultivated groups (3/21 genes in *dicoccum*, 6/20 in *aestivum*, and 8/21 in *durum*). We detected only one instance of departure from the null model of neutral demographic bottleneck. In the *aestivum*

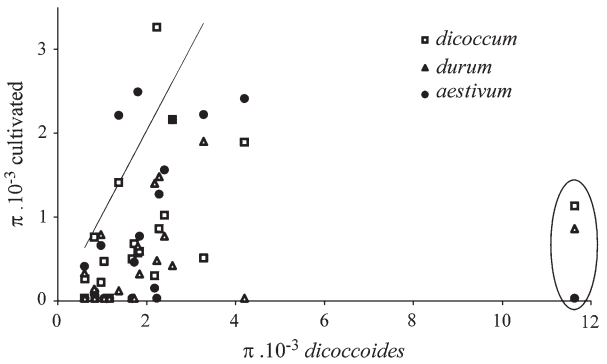

FIG. 5.—Joint patterns of nucleotide diversity observed in wild *dicoccoides* and cultivated relatives at 21 gene fragments, with each gene plotted 3 times. The $\pi_{wild}/\pi_{cultivated}$ ranged from 0 to more than 1, depending on the gene considered. The line (slope = 1) indicates equivalent levels of diversity in *dicoccoides* and the cultivated taxa. The encircled points represent diversity at the *HgA* locus in the 3 cultivated groups.

group, the pattern of *HgA* polymorphism, as summarized by π, cannot be explained by the estimated mean bottleneck intensity of $\alpha_{aestivum}$ = 4.20 (*P* value = 0.0054). Even when the upper limit of the CI of $\alpha_{aestivum}$ (5.74) was used for coalescent simulation, the pattern of polymorphism at *HgA* remained atypical (*P* value = 0.0162). In such multiple testing, with an individual threshold value of 5% for significance, 1 gene of the 20 tested in *aestivum* would be expected to give false-positive results. This weakens the evidence for a possible selective event on *HgA*, which should be viewed with caution.

Discussion

Relationships between Wild Emmer and Cultivated Wheats

Wild emmer has been identified as the wild progenitor of cultivated wheat. Current populations of *dicoccoides* have been reported to fall into 2 genetically different groups (Ozkan et al. 2005). In our *dicoccoides* sample, we detected no population structure without a priori assumption and no significant correlation between genetic and geographic distances on a large scale. However, we detected low, but significant, *F*st values between the 2 groups previously identified (a Western race and a Central-Eastern race). Our sample and data set may have been too small to detect weak genetic differentiation between *dicoccoides* groups. As our *dicoccoides* sample contains accessions from the whole species distribution, we could assume that this sample covers a large proportion of the diversity available in the wild species.

The tree reconstructed from the concatenated 20 gene fragments revealed the distribution of nucleotide diversity within the 4 groups (wild, domesticated, durum, and bread wheat). The diversity in the cultivated group is clearly a subset of the diversity of the wild group, as would be expected for a domestication event (Buckler et al. 2001). The monophyly of all the cultivated individuals in the tree is consistent with a single domestication event for emmer wheat (Zohary 1999). Our results are not consistent with recent

Table 3
Nucleotide Diversity in Wild and Domesticated Relatives

Mating System	Diversity in Wild(10^{-3})	Diversity in Cultivated (10^{-3})	Loci	$L\pi$ (%)	References
	Zea mays ssp. *parviglumis*	*Zea mays* ssp. *mays*			
Outbreeding	$\pi_{total} = 9.7$	$\pi_{total} = 6.4$	774	35	Wright et al. (2005)
	$\pi_{silent} = 21.1$	$\pi_{silent} = 13.1$	12	38	Tenaillon et al. (2004)
	Medicago sativa ssp. *sativa*	*M. s.* ssp. *sativa*	2		Muller et al. (2006)
Outbreeding	$\pi_{total} = 20.2$	$\pi_{total} = 13.5$		31	
	$\pi_{silent} = 29$	$\pi_{silent} = 20$		31	
	Helianthus annuus	*H. annuus*	9		Liu and Burke (2006)
Outbreeding	$\pi_{total} = 12.8$	$\pi_{total} = 5.6$		55	
	$\pi_{silent} = 23.4$	$\pi_{silent} = 9.6$		59	
Mixed	*Pennisetum glaucum*	*P. glaucum*	1		Gaut and Clegg (1993)
	$\theta_{silent} = 3.6$	$\theta_{silent} = 2.4$		33	
	Glycine soja	*Glycine max*	102		Hyten et al. (2006)
Inbreeding	$\pi_{total} = 2.17$	$\pi_{total} = 1.43$		34	
	$\pi_{silent} = 2.76$	$\pi_{silent} = 1.77$		36	
	Hordeum spontaneum	*Hordeum vulgare*			
Inbreeding	$\pi_{silent} = 16.7$	$\pi_{silent} = 7.1$	5	57	Caldwell et al. (2006)
	$\pi_{total} = 8.3$	$\pi_{total} = 3.1$	7	62	Kilian et al. (2006)
	Triticum turgidum ssp. *dicoccoides*	*Triticum turgidum* ssp. *dicoccum*	21		This study
Inbreeding	$\pi_{silent} = 3.6$	$\pi_{silent} = 1.2$		65	
	$\pi_{total} = 2.7$	$\pi_{total} = 0.8$		70	

NOTE.—The loss of diversity during domestication was calculated as $L\pi = 1 - \pi_{dom}/\pi_{wild}$ and nucleotide diversity calculated on silent sites (π_{silent}) and on all sites (π_{total}) are reported (when possible).

suggestions of a possible diphyletic origin for domesticated tetraploid wheats as suggested by Ozkan et al. (2005). The *dicoccum* sequences are widely distributed throughout wheat lineages, spanning the whole range of diversity found in the cultivated group. *Durum* individuals fall into a single clade including some *dicoccum* individuals: these observations are consistent with *dicoccum* being a progenitor of durum wheat. It is not possible to identify precisely which tetraploid donated its A and B genomes to *aestivum*, but *durum* is unlikely the donor subspecies because its nucleotide diversity does not include that of *aestivum*.

Genetic Diversity in *dicoccoides*

The mean nucleotide diversity observed for these 21 genes ($\pi_{total} = 0.0027$ and $\pi_{silent} = 0.0036$) suggests that *dicoccoides* is not a highly polymorphic species. All else being equal, self-fertilizing species are expected to have a lower diversity level than outcrossing species. Selfing reduces effective population size N_e by reducing gamete sampling, and because of low effective recombination rates, hitchhiking effects further reduce diversity (Charlesworth and Wright 2001). Inbreeding and asexual species often have a life history involving frequent local colonization and extinction events, potentially reducing diversity even further (Kimura and Ohta 1971; Charlesworth D and Charlesworth B 1995). *Triticum dicoccoides* displays lower levels of variation than teosinte ($\pi_{total} = 0.0097$, table 3). Another highly inbreeding species, *Glycine soja*, from which soybean was domesticated, also has low levels of diversity ($\pi_{total} = 0.0022$, Hyten et al. 2006). But *dicoccoides* also has a lower level of diversity than the mean observed in a survey of selfing species (mean $\pi_{total} = 0.006$, Glémin et al. 2006). There may be several reasons for this. First, *dicoccoides* arose through a relatively recent allopolyploidy event that may have resulted in a large decrease in diversity in the new species with respect to its diploid ancestors. As the nucleotide mutation rate is low, it is likely to take a long time for diversity to be restored through mutation (Lande and Barrowclough 1987). Thus, the mutation-drift equilibrium may not yet have been reached in *dicoccoides*. The small effective population size of the current population of *dicoccoides* may also account for the low level of diversity. Using microsatellite markers and assuming mutation-drift equilibrium, Thuillet et al. (2005) estimated N_e at 32,500 for *dicoccoides*. The spread of agriculture might have restricted the range of *dicoccoides*, potentially accounting for this low effective population size.

Consequences of Domestication History for DNA Sequences: A Drastic Loss of Diversity

Nucleotide diversity levels were found to be much lower in the 3 cultivated forms than in the wild pool. Assuming that our sample of *dicoccoides* accurately reflects the diversity of the wild progenitor of cultivated wheat 12,000 years ago, initial diversity was reduced by 69% in *aestivum* and 84% in *durum*. Considering the Central East group alone, the diversity reduction associated with domestication is a bit lower ($L\pi_{total} = 67\%$ from Central East population against 70% from the whole wild sample). The increase in Tajima's D from *dicoccoides* ($D = -0.77$, $P < 0.001$) to domesticated wheats ($D = -0.55$, $P = 0.015$; $D = -0.45$, not significant; and $D = 0.48$, $P = 0.041$ for *dicoccum*, *durum*, and *aestivum*, respectively) is also a signature of a recent bottleneck (Tajima 1989), as observed in maize (Wright et al. 2005).

Major losses of neutral diversity have already been demonstrated in the history of *T. turgidum* ssp. by microsatellite analysis (Thuillet et al. 2005). Our coalescent simulations suggest the average domestication bottleneck intensity (from *dicoccoides* to *dicoccum*) of about 3.15,

equivalent to an effective population size of 10,317 in *dicoccum*, assuming that the N_e of *dicoccoides* is 32,500. The nucleotide diversity in bread wheat could be accounted for a bottleneck intensity of 4.2 (corresponding to an N_e of 7,738). Durum wheat experienced the most severe bottleneck (N_e = 5,575), with a population size about one sixth that of wild *dicoccoides*. Using microsatellites, the N_e of domesticated emmer, *dicoccum*, was estimated at 12,000. These 2 estimates of the intensity of the domestication bottleneck in *dicoccum* are qualitatively similar, but the loss of diversity is somewhat greater when estimated with sequence data than with microsatellites, as also reported in a recent study of sunflower domestication (Liu and Burke 2006).

The loss of nucleotide diversity (total and silent) we found during domestication is one of the largest reported so far for a crop species (table 3). Most crops have nucleotide diversities about 30% lower than that of their wild progenitor. However, it is worth noting that wheat and barley lost high and similar amount of silent diversity, 65% and between 57% and 73%, respectively (Caldwell et al. 2006; Kilian et al. 2006).

After domestication, subspecies *durum* and *aestivum* were subject to additional selective events during the evolution of landraces and modern breeding. In durum wheat, 84% of the nucleotide diversity originally present in *dicoccoides* has been lost, with only 20 of the 212 SNPs identified in the wild *dicoccoides* segregating in elite varieties.

Previous studies have reported the existence of at least 2 genetically different progenitors of the D genome of *aestivum*, suggesting independent polyploidization events (Dvorak et al. 1998; Giles and Brown 2006). Nucleotide diversity has been reported to be 30 times higher in *Ae. tauschii* than in the D genome of *T. aestivum* (Caldwell et al. 2004). Thus, if only a few *Ae. tauschii* individuals were involved in the creation of *aestivum*, then only a few tetraploid progenitors are likely to have been involved in the founding of amphiploids. Two groups of *aestivum* (marked as I and II on fig. 2) were identified in the phylogenetic tree reconstructed from nucleotide diversity in the A and B genomes and were confirmed by the STRUCTURE analysis. Positive Tajima's D (D = 0.48357, P = 0.041) also indicates possible population subdivision. These findings support a diphyletic (at least) origin for bread wheat involving genetically different tetraploid progenitors (AABB genome). As *aestivum* is believed to have arisen from rare intergeneric crosses between cultivated tetraploid wheat and the wild diploid *Ae. tauschii*. Thus, it is surprising that *dicoccum* does not include significantly more diversity than *aestivum*. Recurrent gene flow between the tetraploid and hexaploid forms after the emergence of hexaploid forms would have been required to restore the level of diversity of the A and B genomes of *aestivum* after polyploidization. Indirect measurements of sequence polymorphism based on restriction fragment length polymorphism have already suggested the existence of gene flow from parental species to polyploids, especially from *dicoccoides* (Dvorak et al. 2006). As *dicoccum* sequences are widely distributed throughout the tree and span the whole range of diversity found in *aestivum*, we also suggest that gene flow occurred between neighboring *dicoccum* populations.

Using the Demographic Model to Detect Selection

In theory, diversity surveys for identifying selected genes can be applied to any domesticated animal or plant. However, the power of such approaches depends on the relative levels and patterns of diversity for neutral and selected genes in the wild taxon. If neutral genes retain very little diversity after domestication, it is difficult to discriminate neutral from selected genes (Yamasaki et al. 2005). Bottleneck intensity cannot be estimated for nonpolymorphic genes in the cultivated population. However, such genes may be good candidates for selection.

Wright et al. (2005) estimated that 2–4% of maize genes were subject to selection during maize domestication. We found evidence for a similar proportion in wheat domestication, with only 1 of the 21 loci analyzed presenting a pattern of diversity loss suggestive of selection. Although the evidence for possible selection acting on *HgA* should be interpreted with caution, this locus presents a striking pattern of polymorphism (fig. 5). It has been annotated as a hypothetical gene located in the "hardness" locus, about 30 kb from *Gsp1A* (Chantret et al. 2005). *Gsp1A* is thought to be involved in controlling grain softness (Morris 2002), an important trait for end-use quality in wheat. No polymorphism was observed for *Gsp1A* in the *durum* group and for *Gsp1B* in the *aestivum* group, whereas these 2 genes harbor 26 and 7 polymorphic sites, respectively, in the wild *dicoccoides*. The lack of diversity in 1 of the 2 copies in both cultivated wheats suggests that this gene may have been the target of selection during domestication. The *HgA*-linked locus may have been subject to hitchhiking during selection, but further investigations of this candidate region are required to confirm this hypothesis. The authors would like to draw the reader's attention to a recent study of (Luo et al. 2007) where the wild emmer population structure is analyzed on the basis of the restriction fragment length polymorphism at 131 loci.

Supplementary Material

Supplementary tables S1, S2, S3, and S5 and figure S4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Note Added in Proof

The authors would like to draw the reader's attention to a recent study of Luo et al. where the wild emmer population structure is analyzed on the basis of the restriction fragment length polymorphism at 131 loci. (Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor. Appl. Genet. 114:947–959).

Acknowledgments

The authors thank Y. Vigouroux, N. Chantret, S. De Mita, and A. Tsitrone for discussion as well as F. Salamini and 2 anonymous reviewers for helpful comments on an earlier version of the manuscript. This work was supported

by a grant from Tritipol program founded by Institut National de la Recherche Agronomique and the Bureau des Ressources Génétiques.

Literature Cited

Bar-Yosef O. 1998. The Natufian culture in the Levant, threshold to the origins of agriculture. Evol Anthropol. 6:159–177.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004. GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171. Université de Montpellier II, Montpellier.

Buckler E, Thornsberry JM, Kresovich S. 2001. Molecular diversity, structure and domestication of grasses. Genet Res. 77:213–218.

Caldwell KS, Dvorak J, Lagudah ES, Akhunov E, Luo MC, Wolters P, Powell W. 2004. Sequence polymorphism in polyploid wheat and their d-genome diploid ancestor. Genetics. 167:941–947.

Caldwell KS, Russell J, Langridge P, Powell W. 2006. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. Genetics. 172: 557–567.

Chantret N, Salse J, Sabot F, et al. (19 co-authors). 2005. Molecular basis of evolutionary events that shaped the *hardness* locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell. 17:1033–1045.

Charlesworth D, Charlesworth B. 1995. Quantitative genetics in plants—the effect of the breeding system on genetic variability. Evolution. 49:911–920.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. Curr Opin Genet Dev. 11:685–690.

Daud HM, Gustafson JP. 1996. Molecular evidence for Triticum speltoides as a B-genome progenitor of wheat (*Triticum aestivum*). Genome. 39:543–548.

David JL, Bataillon T, Poirier S, Roumet P, Santoni S, Thuillet A-C. 2003. Impact of demographic and selective events on the current genetic diversity of the *Triticum turgidum* complex. In: Pogna NE, Romano M, Pogna EA, Galterio G, editors. Tenth International Wheat Genetics Symposium, S.I.M.I., Paestum, Italy. p. 7–10.

Dvorak J, Akhunov ED. 2005. Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. Genetics. 171:323–332.

Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC. 2006. Molecular characterization of a diagnostic DNA marker for domesticated tetraploid wheat provides evidence for gene flow from wild tetraploid wheat to hexaploid wheat. Mol Biol Evol. 23:1386–1396.

Dvorak J, Diterlizzi P, Zhang HB, Resta P. 1993. The evolution of polyploid wheats—identification of the A-genome donor species. Genome. 36:21–31.

Dvorak J, Luo M-C, Yang Z-L, Zhang H-B. 1998. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet. 97:657–670.

Dvorak J, Zhang HB. 1990. Variation in repeated nucleotide sequences sheds light on the phylogeny of the wheat B and G genomes. Proc Natl Acad Sci USA. 87:9640–9644.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS. 1998. Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA. 95:4441–4446.

Felsenstein J. 2005. PHYLIP (phylogeny inference package) version 3.6. Seattle (WA): Department of Genome Sciences, University of Washington.

Gaut BS, Clegg MT. 1993. Nucleotide polymorphism in the Adh1 locus of pearl-millet (Pennisetum-Glaucum) (Poaceae). Genetics. 135:1091–1097.

Giles RJ, Brown TA. 2006. GluDy allele variations in *Aegilops tauschii* and *Triticum aestivum*: implications for the origins of hexaploid wheats. Theor Appl Genet. 112:1563–1572.

Glémin S, Bazin E, Charlesworth D. 2006. Impact of mating systems on patterns of sequence polymorphism in flowering plants. Proc R Soc Lond B Biol Sci. 273:3011–3019.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Harlan JR. 1992. Crops and man. Foundations for Modern Crop Science Series. Madison (WI): American Society of Agronomy and Crop Science Society of America.

Heun M, SchaferPregl R, Klawan D, Castagna R, Accerbi M, Borghi B, Salamini F. 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. Science. 278:1312–1314.

Huang S, Sirikhachornkit A, Su XJ, Faris J, Gill B, Haselkorn R, Gornicki P. 2002. Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA. 99:8133–8138.

Hudson RR. 1990. Gene genealogies and the coalescent process. In: Futuyma DJ, Antonovics J, editors. Oxford surveys in evolutionary biology. New York: Oxford University Press. p. 1–44.

Hudson RR. 2001. Two-locus sampling distributions and their application. Genetics. 159:1805–1817.

Hudson RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics. 18: 337–338.

Hyten DL, Song Q, Zhu Y, Choi I-Y, Nelson RL, Costa JM, Specht JE, Shoemaker RC, Cregan PB. 2006. Impacts of genetic bottlenecks on soybean genome diversity. Proc Natl Acad Sci USA. 103:16666–16671.

Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B. 2005. Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res. 15:526–536.

Khlestkina EK, Salina EA. 2001. Genome-specific markers of tetraploid wheats and their putative diploid progenitor species. Plant Breed. 120:227–232.

Kihara H. 1944. Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agric Hortic (Tokyo). 19: 13–14.

Kilian B, Ozkan H, Kohl J, von Haeseler A, Barale F, Deusch O, Brandolini A, Yucel C, Martin W, Salamini F. 2006. Haplotype structure at seven barley genes: relevance to gene pool bottlenecks, phylogeny of ear type and site of barley domestication. Mol Genet Genomics. 276:230–241.

Kimura M, Ohta T. 1971. Theoretical topics in population genetics. Princeton (NJ): Princeton University Press.

Lande R, Barrowclough GF. 1987. Effective population size, genetic variation, and their use in population management. Viable populations for conservation. Cambridge: Cambridge University Press. p. 87–123.

Lev-Yadun S, Gopher A, Abbo S. 2000. Archaeology—the cradle of agriculture. Science. 288:1602–1603.

Liu A, Burke JM. 2006. Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics. 173:321–330.

Luo M-C, Yang Z-L, You FM, Kawahara T, Waines JG, Dvorak J. 2007. The structure of wild and domesticated emmer wheat populations, gene flow between them, and the site of emmer domestication. Theor Appl Genet. 114:947–959.

McFadden ES, Sears ER. 1946. The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J Hered. 37: 81–89.

Mori N, Liu YG, Tsunewaki K. 1995. Wheat Phylogeny Determined by Rflp Analysis of Nuclear-DNA. 2. Wild Tetraploid Wheats. Theor Appl Genet. 90:129–134.

Mori N. 2003. Wheat domestication: when, where and how? Plant Cell Physiol. 44:S2–S2.

Morris CF. 2002. Puroindolines: the molecular genetic basis of wheat grain hardness. Plant Mol Biol. 48:633–647.

Muller M-H, Poncet C, Prosperi JM, Santoni S, Ronfort J. 2006. Domestication history in the *Medicago sativa* species complex: inferences from nuclear sequence polymorphism. Mol Ecol. 15:1589–1602.

Nesbitt M, Samuel D. 1996. From staple crop extinction? The archaeology and history of the hulled wheats. In: Padulosi S, Hammer K, Heller J, editors. Hulled wheat. Promoting the conservation and use of underutilized and neglected crops. Rome (Italy): International Plant Genetic Resources Institute. p. 41–100.

Nesbitt M, Samuel D. 1998. Wheat domestication: archaeobotanical evidence. Science. 279:1433–1433.

Nordborg M. 2003. Coalescent theory. In: Balding D, Bishop M, Cannings C, editors. Handbook of statistical genetics. Chichester (UK): Wiley. p. 602–635.

Ozkan H, Brandolini A, Pozzi C, Effgen S, Wunder J, Salamini F. 2005. A reconsideration of the domestication geography of tetraploid wheats. Theor Appl Genet. 110: 1052–1060.

Ozkan H, Brandolini A, Schafer-Pregl R, Salamini F. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. Mol Biol Evol. 19:1797–1801.

Peakall R, Smouse PE. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes. 6:288–295.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

Qi LL, Echalier B, Chao S, et al. (47 co-authors). 2004. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics. 168:701–712.

Ravel C, Praud S, Murigneux A, et al. (12 co-authors). 2006. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L). Genome. 49: 1131–1139.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R, Martin W. 2002. Genetics and geography of wild cereal domestication in the near east. Nat Genet. 3:429–441.

Sasanuma T, Chabane K, Endo TR, Valkoun J. 2002. Genetic diversity of wheat wild relatives in the near east detected by AFLP. Euphytica. 127:81–93.

Septiningsih EM, Prasetiyono J, Lubis E, Tai TH, Tjubaryat T, Moeljopawiro S, McCouch SR. 2003. Identification of quantitative trait loci for yield and yield components in an advanced backcross population derived from the *Oryza sativa* variety IR64 and the wild relative O-rufipogon. Theor Appl Genet. 107:1419–1432.

Sourdille P, Perretant MR, Charmet G, Leroy P, Gautier MF, Joudrier P, Nelson JC, Sorrells ME, Bernard M. 1996. Linkage between RSLP markers and genes affecting kernel hardness in wheat. Theor Appl Genet. 93:580–586.

Staden R, Judge DP, Bonfield JK. 2001. Sequence assembly and finishing methods. Methods Biochem Anal. 43:303–322.

Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.

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

Tanksley SD, McCouch SR. 1997. Seed banks and molecular maps: unlocking genetic potential from the wild. Science. 277:1063–1066.

Tanno K, Willcox G. 2006. How fast was wild wheat domesticated? Science. 311:1886–1886.

Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS. 2004. Selection versus demography: a multilocus investigation of the domestication process in maize. Mol Biol Evol. 21:1214–1225.

Thuillet A-C, Bataillon T, Poirier S, Santoni S, David JL. 2005. Estimation of long-term effective population sizes through the history of durum wheat using microsatellite data. Genetics. 169:1589–1599.

Vigouroux Y, Jaqueth JS, Matsuoka Y, Smith OS, Beavis WD, Smith JSC, Doebley J. 2002. Rate and pattern of mutation at microsatellite loci in maize. Mol Biol Evol. 19:1251–1260.

Wakeley J, Hey J. 1997. Estimating ancestral population parameters. Genetics. 145:847–855.

Weiss G, von Haeseler A. 1998. Inference of population history using a likelihood approach. Genetics. 149:1539–1546.

Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. 2005. The effects of artificial selection on the maize genome. Science. 308:1310–1314.

Yamasaki M, Tenaillon MI, Bi IV, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell. 17:2859–2872.

Zohary D. 1999. Monophyletic vs. polyphyletic origin of the crops on which agriculture was founded in the near east. Genet Resour Crop Evol. 46:133–142.

Zohary D, Hopf M. 2000. Domestication of plants in the Old World: the origin and spread of cultivated plants in West Asia Europe and the Nile Valley. New York: Oxford University Press.

William Martin, Associate Editor

Accepted March 27, 2007