

Estimation of Long-Term Effective Population Sizes Through the History of Durum Wheat Using Microsatellite Data

A.-C. Thuillet, T. Bataillon, S. Poirier, S. Santoni and J. L. David¹

Institut National de la Recherche Agronomique-UMR 1097 Diversité et Génomes des Plantes Cultivées, 34130 Mauguio, France

Manuscript received April 6, 2004

Accepted for publication November 13, 2004

ABSTRACT

Estimation of long-term effective population size (N_e) from polymorphism data alone requires an independent knowledge of mutation rate. Microsatellites provide the opportunity to estimate N_e because their high mutation rate can be estimated from observed mutations. We used this property to estimate N_e in allotetraploid wheat *Triticum turgidum* at four stages of its history since its domestication. We estimated the mutation rate of 30 microsatellite loci. Allele-specific mutation rates μ were predicted from the number of repeats of the alleles. Effective population sizes were calculated from the diversity parameter $\theta = 4N_e\mu$. We demonstrated from simulations that the unbiased estimator of θ based on Nei's heterozygosity is the most appropriate for estimating N_e because of a small variance and a relative robustness to variations in the mutation model compared to other estimators. We found a N_e of 32,500 individuals with a 95% confidence interval of [20,739; 45,991] in the wild ancestor of wheat, 12,000 ([5790; 19,300]) in the domesticated form, 6000 ([2831; 9556]) in landraces, and 1300 ([689; 2031]) in recent improved varieties. This decrease illustrates the successive bottlenecks in durum wheat. No selective effect was detected on our loci, despite a complete loss of polymorphism for two of them.

THE effective population size (N_e) is defined as the size of an ideal population that would present the same amount of genetic drift as the population under study. Effective size determines the rate at which populations lose alleles through random genetic drift and, together with the mutation rate, it determines the number of alleles expected in populations. This parameter is therefore a measure of the genetic diversity that can be maintained in populations. Demographic events (expansions or reductions) can have dramatic effects on the long-term effective size of a population. Domestication and subsequent agronomic improvement of crop species are usually accompanied by such changes in population size. Population size reductions, typically referred to as genetic bottlenecks, have usually led to decreased levels of diversity in cultivated crops relative to that observed in their wild progenitors (see BUCKLER *et al.* 2001 for a review of grasses). The management of genetic resources for cultivated plant species would benefit from a better understanding of how genetic diversity has changed throughout their evolutionary his-

tory. In this context, knowledge of how effective population sizes have changed in crop species is of special interest, in that N_e is expected to reflect the reduction of genetic diversity that has occurred at different stages in the history of crop domestication.

The section of allotetraploid wheats with AB genomes (*Triticum turgidum* $2n = 4x = 28$) is an appropriate model to study the impact of demographic events on effective population size since the history of wheat domestication is well documented and the different gene pools (*e.g.*, wild and domesticated) are still available as accessions conserved in seed banks. Durum wheat (*T. turgidum* ssp. *durum*) is traditionally grown around the Mediterranean Sea and is currently the most common cultivated form of allotetraploid wheat. Wheat domestication took place $\sim 12,000$ years ago in the Near East, with the wild ancestor (*T. turgidum* ssp. *dicoccoides*) giving rise to the first domesticated form (emmer wheat, *T. turgidum* ssp. *dicoccum*; ZOHARY and HOPF 2000). About 2000 years after this event, the spread of agriculture from this region led to the expansion of *T. turgidum* ssp. *dicoccum* (mainly through Europe). During the same period, durum wheat (*T. turgidum* ssp. *durum*), which is characterized by free threshing, appeared in the Near East and replaced its ancestor *T. turgidum* ssp. *dicoccum* to become the major cultivated form of allotetraploid wheat by the second millennium BC (MAIER 1996; NESBITT and SAMUEL 1996; ZOHARY and HOPF 2000). The most recent history of durum wheat has been marked by modern genetic improvement, involving the replacement of landraces by inbred varieties and the introduc-

Sequence data from this article have been deposited with the Gen Bank Data Libraries under accession nos. AF275895, AF275896, AF275897, AF275898, AF275899, AF275900, AF275901, AY579587, AY579588, AY579589, AY579590, AY579591, AY579592, AY579593, AY579594, AY579595, AY579596, AY579597, AY579598, AY579599, AY579600, AY579601, AY579602, AY579603, AY579604, AY579605, AY579606, and AY579607.

¹Corresponding author: INRA-UMR Diversité et Génomes des Plantes Cultivées, Domaine de Melgueil, 34130 Mauguio, France.
E-mail: david@ensam.inra.fr

tion of dwarfing genes (second part of the 20th century). These historical events are likely to have altered the genetic diversity of wheat through two mechanisms: first, a genome-wide loss of diversity is expected due to the sampling of a few individuals at each historical transition (a bottleneck effect); second, in genomic regions harboring genes coding for favorable traits, even more severe losses of diversity would be expected as a consequence of direct selection on these genes (see, for instance, WANG *et al.* 1999 for a striking example on the domestication of maize and its effects on the diversity of the *tb1* locus).

Here we seek to estimate the long-term effective population size of gene pools representing several key stages of the history of durum wheat, using contemporary samples. Estimates of N_e are notoriously difficult to obtain from samples at a single point in time and usually only the product $N_e\mu$ is estimable, where N_e is the effective population size and μ the mutation rate of the locus used for surveying diversity in the sample. However, if independent knowledge of the mutation rate is available, it is possible to estimate N_e from observed polymorphism (NEI and GRAUR 1984). Microsatellite markers present high mutation rates that can therefore be estimated from direct observation of mutation events (ELLEGREN 2000a; XU *et al.* 2000; THUILLET *et al.* 2002). Estimates of N_e based on these markers have been obtained for several species (LEHMANN *et al.* 1998; SPONG *et al.* 2000; PASCUAL *et al.* 2001; VIGOUROUX *et al.* 2002). Over the last decade, a large collection of data has been generated concerning microsatellite mutation rates and their pattern of evolution. In particular, microsatellite allele length has been shown to be positively correlated with the rate of mutation (BRINKMANN *et al.* 1998; SCHLÖTTERER *et al.* 1998; VIGOUROUX *et al.* 2002; BRODEHE *et al.* 2004). Making use of this information should allow one to obtain increased accuracy of estimating N_e values using these markers (ELLEGREN 2000b; SCHLÖTTERER 2000; XU *et al.* 2000).

In this article, we first provide direct estimates of mutation rates for a set of microsatellite loci in *T. t. durum* and develop a calibration allowing the prediction of allele-specific mutation rates from their allele length. Second, using these calibrated microsatellites to survey diversity, we estimate absolute values of the effective population sizes in four taxonomic groups representative of key stages in the evolution of *T. t. durum*. This allows us to quantify the impact of demographic events on the genetic diversity through the history of durum wheat since its domestication. We also examine severe losses of genetic diversity at several loci in the context of selection for agronomic traits.

MATERIALS AND METHODS

Plant material: *Mutation-accumulation design:* Five hundred inbred lines diverged for 8 or 10 generations of controlled

selfing from a single doubled haploid of the variety *Sham I* were genotyped to obtain microsatellite loci mutation rate estimates. This variety belongs to the subspecies *T. t. durum*. Plant material is reported in detail in THUILLET *et al.* (2002).

Plant material used for the diversity survey: Three hundred fourteen genotypes were sampled to examine diversity at four main stages of the history of the allotetraploid wheat *T. turgidum*. The four groups were defined as follows: 59 accessions originating from the Near East represented the wild emmer wheat (*T. turgidum* ssp. *dicoccoides*); 94 individuals coming from the Near East, Turkey, the Balkans, eastern Europe, Italy, Spain, Russia, and Morocco (accessions were kindly provided by the U.S. Department of Agriculture, International Center for Agricultural Research in the Dry Areas, and International Maize and Wheat Improvement Center) constituted the sample of domesticated emmer wheat (*T. turgidum* ssp. *dicoccum*); 40 accessions, conserved at INRA (Montpellier) were old varieties of durum wheat that were cultivated in the first part of the 20th century; and 121 accessions were French modern varieties of durum wheat registered after 1960, the date corresponding to the introduction of dwarfing genes to produce elite varieties (provided by the Group of Study and Control of Varieties and Seeds, France). We further refer to these four groups as the wild group, domesticated emmer, old durum varieties, and elite durum varieties, respectively.

DNA extraction and microsatellite genotyping: DNA of the mutation-accumulation lines and of inbred lines comprising the four diversity groups was extracted from fresh leaves following a protocol adapted from TAI and TANKSLEY (1990). Amplification reactions for microsatellite genotyping were performed in a final volume of 20 μ l in the presence of 20 ng of template DNA, 1 pmol of the reverse primer, and 4 pmol of the forward primer; 0.2 mM of each deoxynucleotide; 1.5 mM $MgCl_2$; and 1 unit Taq polymerase (Sigma, St. Louis). The forward primer was 5'-labeled with one of the three fluorophores (6FAM, NED, or HEX). PCR was carried out using a PTC 100 thermocycler (MJ Research, Watertown, MA). After 5 min at 94°, 30 cycles were performed for 30 sec at 94°; 30 sec at 50, 55, or 60° (depending on the locus); and 30 sec at 72°, followed by a final extension step of 5 min at 72°. Amplified products were detected on an ABI (Columbia, MD) prism 3100 Genetic Analyzer. Samples were prepared by adding 3 μ l of diluted PCR products to 6.875 μ l formamide and 0.125 μ l GenSize 500 Rox. Analysis was performed using the Gene Scan 3.1 and Genotyper 2.5 software (Applied Biosystems, Foster City, CA).

Direct estimation of the mutation rates: Mutation rates for 30 microsatellite loci, mapped on the A and B genomes of the bread wheat *T. aestivum* (RÖDER *et al.* 1998a), were estimated directly in the mutation-accumulation lines derived from the variety *Sham I*. Seven loci were scored after 8 generations (THUILLET *et al.* 2002) and 23 loci were scored after 10 generations (3 loci reported in THUILLET *et al.* 2002; 20 loci in this study; see Table 1). Being allotetraploid, segregation of alleles at a given locus is disomic. Therefore, at each locus, the mutation rate was estimated as the number of mutant alleles observed in the 500 mutation-accumulation lines divided by the number (n) of possible mutation events scored ($n = L(2 - 2^{1-g} + g)$), where L is the number of lines screened and g the number of generations considered, $n = 4996$ for 500 lines and 8 generations, and $n = 5999$ for 500 lines and 10 generations; see THUILLET *et al.* (2002) for details. Exact confidence intervals around these estimates were computed by assuming that the total number of mutation events detected at a given locus is Poisson distributed (CASELLA and BERGER 1990).

Influence of the allele length on the mutation rate of microsatellite alleles: The number of repeats of the initial allele was obtained after cloning and sequencing PCR products obtained

on *Sham I* (see THUILLET *et al.* 2002). The relationship between the mutation rate of an allele ($\hat{\mu}$) and its number of repeats (R) was investigated by a regression analysis in the *Sham I* mutation-accumulation design. For imperfect loci, we considered each uninterrupted part of the sequence as an independent perfect repeated array. Thus, R was set as the arithmetic mean of all uninterrupted parts of the repeated sequence, and the observed corresponding mutation rate at the locus was divided by the number of uninterrupted parts of the sequence. This aims to take into account the fact that interrupted microsatellites seem to present lower mutation rates than perfect microsatellites for a similar total length (BRINKMANN *et al.* 1998). Loci for which no mutant allele was observed were not used in the regression since no estimate of the mutation rate was available in this case. Indeed, single estimates are needed for every locus to build the regression and using zero as an estimate for these loci would bias the equation of the regression line. Therefore we discarded from our regression analysis loci for which no estimate of the mutation rate is available. These loci were used instead to cross-validate the equation obtained from the regression analysis $\hat{\mu} = f(R)$. If the prediction of their mutation rate based on their number of repeats is valid, the predicted value should lie within the 95% confidence interval obtained for these loci (based on zero mutation observed).

Estimation of locus-specific mutation rates in each group of *T. turgidum*: The four groups of *T. turgidum* were genotyped for a subset of 15 loci among the 30 loci analyzed on the mutation-accumulation lines. The equation $\hat{\mu} = f(R)$ deduced from the regression analysis (see RESULTS below) was used to infer allele-specific mutation rates, at each locus in each group. The number of repeats of each individual allele was deduced from its length in base pairs compared to the length of the (sequenced) reference allele in *Sham I*. As differences in allele size among alleles consisted of even numbers of bases, we assumed that the flanking sequence did not vary among alleles and that all the differences in size we detected were due to variations in the number of motifs in the repeated array. Mutation rates were calculated for each allele and we tested the homogeneity of the predicted mutation rates between loci and among groups through an analysis of variance (SAS 1999–2000).

Estimation of θ in the groups of *T. turgidum*: $\theta = 4N_c\mu$ can be estimated (i) from Nei's heterozygosity (H_c ; NEI 1987) or from the variance of allele size under a stepwise mutation model (SMM, where mutation events involve a single repeat) and (ii) from H_c or from the number of alleles under an infinite allele model (IAM, where each mutation event creates a new allele not present in the population). Most of our loci are compatible with a SMM assumption (see below), which is moreover reasonable for microsatellites as for these loci alleles can mutate to a state that is already preexistent in the population. The estimator of θ based on the variance of allele size has a high sampling variance (KIMMEL and CHAKRABORTY 1996) and is sensitive to mutation events involving more than one repeat, since these events would increase the variance in allele size dramatically. We consequently focus here on Nei's heterozygosity under a SMM. H_c was calculated at the 15 loci in the four groups. Under the assumptions of a SMM and of mutation-drift equilibrium, the parameter $\theta = 4N_c\mu$ is estimated from H_c as $\hat{\theta}_{H_c} = 1/2[(1 - H_c)^{-2} - 1]$ (OHTA and KIMURA 1973). This estimator being biased (ZOUROS 1979), we used the correction proposed by XU and FU (2004) and refer to the bias-corrected estimator of θ as $\hat{\theta}_F$. We estimated the effective population size of group i and at locus j ($\hat{N}_{c_{ij}}$) by dividing $\hat{\theta}_{F_{ij}}$ by four times the average predicted mutation rate at locus j in group i , $\hat{\mu}_{ij}$, i.e., $\hat{N}_{c_{ij}} = \hat{\theta}_{F_{ij}}/4\hat{\mu}_{ij}$. For each group i , we averaged over the j loci the values of $\hat{N}_{c_{ij}}$. A bootstrap

(10,000 resampling) over loci was performed to obtain an $\sim 95\%$ confidence interval around the effective population size in each group. Loci departing significantly from neutrality and from the SMM were not used in the estimation of N_c (see the following sections).

Computer simulations: Under a SMM and under the mutation-drift equilibrium assumption, $\hat{\theta}_F$ is better than other existing estimators of θ for microsatellite data, since it is unbiased and has the smallest variance among currently available estimators (XU and FU 2004). However, XU and FU (2004) also showed that this estimator can be dramatically misleading when mutation events involve more than one repeat. Moreover, in our case populations have been sampled to represent postbottleneck stages: if populations have not recovered an equilibrium state, this might also affect the value of $\hat{\theta}_F$. Therefore we performed computer simulations to investigate the behavior of $\hat{\theta}_F$ under a two-phase model (TPM), where mutation events involving more than one repeat gain/loss can occur, (DI RIENZO *et al.* 1994) and after a bottleneck. In addition, we tested (i) loci for a deviation from the SMM and (ii) groups for signature of strong demographic nonequilibrium (see the following section).

Simulations were based on the coalescent process following the algorithm proposed by HUDSON (1990) to simulate population samples. Under each mutation model, assuming constant population size, we simulated 1000 samples consisting of 50 genes for 10 values of θ ranging from 1 to 40. The parameters for the TPM were a proportion $p = 0.20$ of multistep mutations where the number of steps was sampled in a geometric distribution of parameter $\alpha = 0.85$, corresponding to a variance $\sigma^2 = 38$. The values chosen here for p and α are consistent with values proposed in the literature to fit actual microsatellite polymorphism data (DI RIENZO *et al.* 1994). The bottleneck was simulated with a sudden reduction of the effective population size as suggested by HUDSON (1990). Through the simulated bottleneck the population size was reduced to one-twentieth of its initial value 10,000 generations before present. After a bottleneck, values of θ ranged from 0.2 to 16.

Detection of loci departing from the SMM: At a microsatellite locus presenting a given number of alleles, some values of the variance of allele size are unlikely under a SMM whereas they are still plausible under a TPM (THUILLET *et al.* 2004). Simulations of a large number of data sets under a broad range of θ (from 1 to 50; THUILLET *et al.* 2004) were used to obtain the joint distribution of the number of alleles (k) and the variance of allele size (V) under a SMM. Ninety-nine percent of the simulated data sets were kept to define a threshold beyond which values of V at a given locus are highly unlikely under the SMM assumption. This test is described in more detail in THUILLET *et al.* (2004) and was applied here in each group of wheat to discard loci likely to evolve in a TPM manner.

Detection of the signature of a past bottleneck in each group: Except for the wild ancestor, the different groups have been sampled to represent the diversity of groups that are likely to have undergone a bottleneck during their history. The assumption underlying our estimation of N_c in each group is that a mutation-drift equilibrium state has been recovered by the population since the bottleneck. If it is not the case, the estimation of θ can be affected. We tested if the signature of a bottleneck was still detectable in each population, including the wild ancestor, since a demographic decrease might have occurred after the domestication event. The test relies on the fact that in a recently bottlenecked population, the number of alleles is reduced faster than the gene diversity (NEI *et al.* 1975). The software Bottleneck was used (CORNUET and LUKART 1996). A Wilcoxon signed-rank test was used to test if a significantly high number of loci showed a heterozygosity

excess or deficiency relative to the expected heterozygosity computed under the equilibrium hypothesis from the number of alleles. Loci suspected to follow a TPM were excluded. Note, however, that among the remaining loci, some might still evolve under a TPM but may be compatible with a SMM if the population has experienced a recent bottleneck (THUILLET *et al.* 2004). Both the SMM and the TPM hypotheses have been considered to compute expected heterozygosities in the different groups at the different loci given the observed number of alleles. The parameters of the TPM were set as previously for simulations ($p = 0.2$, $\sigma^2 = 38$).

Detection of loci departing from the neutral expectation:

To avoid a selective bias in the estimation of effective sizes due to outlier loci and to detect a signature of selection in our data set, we tested each locus used in the diversity survey for a departure from neutral evolution. At a given microsatellite locus, assuming that the mutation rate is identical in two populations, the expectation of the ratio of $\theta = 4N_e\mu$ for a pair of populations estimates the ratio of N_e between both populations. The logarithm of this ratio, using $\hat{\theta}_{H_c}$ to estimate θ , has been shown to follow a normal distribution under neutral evolution (SCHLÖTTERER 2002; KAUER *et al.* 2003). As it appeared from simulation results that the estimation of the ratio calculated from $\hat{\theta}_F$ presented a larger variance than that calculated from $\hat{\theta}_{H_c}$ (data not shown), we kept $\hat{\theta}_{H_c}$ for the calculation of the ratio. The normality of the distribution of the ratio holds when the pattern of evolution does not vary for a given locus from one population to the other, but the distribution of the estimated ratio can depart from normality when a bottleneck occurs, especially for TPM loci (KAUER *et al.* 2003). To account for possible differences in the mutation rates in the different populations, we corrected each ratio by the corresponding group mutation rates in each population. This new ratio, noted $\ln(R[\theta/\mu])$, was calculated for each locus and for three pairs of wheat groups: (a) the wild group and the domesticated group, (b) the domesticated group and the old varieties, and (c) the old varieties and modern varieties. To test departure from neutrality, the expected distribution of the statistic $\ln(R[\theta/\mu])$ was obtained by computer simulations.

The average ratio of N_e between two populations (excluding TPM loci) provided an estimate of the expected reduction of N_e between each pair of populations, which was used to parameterize coalescence simulations and obtain the expected distribution of $\ln(R[\theta/\mu])$ from 1000 coalescent simulations. For each pairwise comparison, loci outside the interval spanning 95% of the expected distribution of $\ln(R[\theta/\mu])$ were considered as outliers potentially affected by selection.

RESULTS

Direct estimation of mutation rates: Loci properties are summarized in Table 1. All loci were dinucleotide repeats, 21 were perfect (*i.e.*, presented an uninterrupted repetition of the microsatellite motif), and 9 were imperfect (*i.e.*, presented an interruption in the repetition of the microsatellite motif). Sequences are available from GenBank under accession nos. AF275895–AF275901 and AY579587–AY579607 for 28 loci. For the two last loci (*Xgwm537* and *Xgwm99*) we could not obtain a clean sequence after several attempts because of technical problems. However, in these two cases, the number of repeats could be deduced approximately, with an uncertainty of plus or minus one repeat. Primers used to amplify loci *Xgwm160* and *Xgwm6* provided two ampli-

cons in *Sham I* (the amplified loci are noted hereafter loci *Xgwm160a*, *Xgwm160b*, *Xgwm6a*, and *Xgwm6b*). The number of repeats (CA or GA) was variable across loci. The mean mutation rate over the 30 loci was 3.1×10^{-4} when a value of zero was considered for loci presenting no mutation. A total of 49 mutation events were observed among 13 loci and the highest observed mutation rate was 2.1×10^{-3} (Table 1). A great majority of gains of one repeat occurred (26 gains against five losses; $\chi^2 = 14.5$, d.f. = 1, $P < 0.0005$). Multistep mutations of two, four, five, or six repeated units also occurred and consisted in losses of repeats except for one gain of two repeats (*Xgwm6a*). The number of repeats deduced from the sequence in *Sham I* was compared to the number of repeats of the sequence at the same locus in the genotype of hexaploid wheat *Chinese spring* (Table 1): no difference of structure was detected for 20 loci among 26 for which the comparison was possible (*i.e.*, perfect or imperfect loci in both species). The conservation of structure of the loci seems consequently a safe assumption within the species *T. turgidum*.

Prediction of mutation rates from the number of repeats: The correlation between the mutation rate (μ) and the number of repeats of the allele (R) was based on the 13 loci for which mutations were observed. When all 13 loci were considered, variation in R explained only 34% of the variation in μ ($P < 0.04$). However, this low value seemed to be due to one particular locus (*Xgwm6a*) that exhibited a particularly high mutation rate despite a low number of repeats. Removing this locus from the analysis resulted in a dramatic increase of the regression coefficient ($R^2 = 0.83$, $P < 0.0001$) and the relationship between allele length (R) and mutation rate (μ) was

$$\mu = 0.00003R - 0.0001. \quad (1)$$

This regression was based on 10 perfect and 2 imperfect loci. Considering the 10 perfect microsatellites alone, the correlation was slightly improved, with a better coefficient of determination ($R^2 = 0.91$, $P < 0.0001$). Equation 1 was kept for further prediction of mutation rates.

At all 6 imperfect loci, and at 7 of 10 perfect loci for which no mutation events were detected in the mutation-accumulation design, the predicted mutation rates in *Sham I* were compatible with the probability of observing zero mutation at the 95% confidence level. Three perfect loci (*Xgwm213*, *Xgwm99*, and *Xgwm372*) had higher predicted mutation rates than the maximum rate expected at the 95% confidence level given zero mutation observed.

Equation 1 was considered robust enough to predict mutation rates in the different groups of wheat. Among the 15 loci that were used to survey microsatellite diversity, 6 were part of the 12 loci used to build the regression analysis. For those 6 loci, the mutation rate predicted by Equation 1 is by construction in agreement with the observed mutation rate in the mutation-accumulation de-

TABLE 1
Direct estimation of mutation rates for 30 microsatellite loci

Locus	Ch	Motif		Mutation rate	C.I. 95%	+2	+1	-1	-2	-4	-5	-6
		<i>Chinese spring</i>	<i>Sham I</i>									
<i>Xgwm164</i>	1A	(GA) ₁₆	(GA) ₂₀	0	0; 5.2×10^{-4}
<i>Xgwm99^a</i>	1A	(GT) ₂₀	(GT) ₂₉	0	0; 5.3×10^{-4}
<i>Xgwm413</i>	1B	(GA) ₁₈	(GA) ₂₀	3.5×10^{-4}	1.1×10^{-4} ; 1.3×10^{-3}	.	.	2
<i>Xgwm153^a</i>	1B	(GA) ₁₈	(GA) _{>34}	10^{-3}	4.7×10^{-4} ; 2.5×10^{-3}	.	4	1
<i>Xgwm339</i>	2A	(GA) ₂₂	(GA) ₂₅	5.3×10^{-4}	1.9×10^{-4} ; 1.6×10^{-3}	.	3
<i>Xgwm372^a</i>	2A	(GA) _{>51}	(GA) ₂₆	0	0; 6.4×10^{-4}
<i>Xgwm312^a</i>	2A	(GA) ₃₇	(GA) ₃₀	6.4×10^{-4}	2.3×10^{-4} ; 1.9×10^{-3}	.	3
<i>Xgwm257^a</i>	2B	(GT) ₃₀	(GT) ₆ GC(GT) ₅	0	0; 6.4×10^{-4}
<i>Xgwm374^a</i>	2B	(GT) ₁₇	(GT) ₁₀	0	0; 6.4×10^{-4}
<i>Xgwm148^a</i>	2B	(GT) ₂₂	(GT) ₁₁	0	0; 6.4×10^{-4}
<i>Xgwm526^a</i>	2B	(GA) ₁₆	(GA) ₁₈	0	0; 6.4×10^{-4}
<i>Xgwm2</i>	3A	(GT) ₁₈	(GT) ₅ GA(GT) ₅	3.6×10^{-4}	1.1×10^{-4} ; 1.3×10^{-3}	.	2
<i>Xgwm247</i>	3B	(GA) ₂₄	(GA) ₁₄	1.7×10^{-4}	4.2×10^{-5} ; 9.6×10^{-4}	.	1
<i>Xgwm285</i>	3B	(GA) ₂₇	(GA) ₁₉	0	0; 5.2×10^{-4}
<i>Xgwm160a</i>	4A	—	(GA) ₁₂ GG(GA) ₇	0	0; 5.3×10^{-4}
<i>Xgwm160b</i>	4A	—	(GA) ₄₊₁₄₊₇₊₂	0	0; 5.2×10^{-4}
<i>Xgwm601</i>	4A	(GA) ₁₇	(GA) ₁₈	0	0; 5.2×10^{-4}
<i>Xgwm6a</i>	4B	—	(GA) ₂₂ AA(GA) ₁₀	2.1×10^{-3}	1.2×10^{-3} ; 3.6×10^{-3}	1	10	.	.	1	.	.
<i>Xgwm6b</i>	4B	—	(GA) ₅₃	1.6×10^{-3}	8.3×10^{-4} ; 2.9×10^{-3}	.	5	2	.	.	1	1
<i>Xgwm495^a</i>	4B	(GA) ₂₀	(GA) ₁₂ A(GA) ₁₂	0	0; 5.3×10^{-4}
<i>Xgwm126</i>	5A	(GT) ₁₅	(GT) ₁₂	—	—
<i>Xgwm156</i>	5A	(GT) ₁₄	(GT) ₁₇	0	0; 5.2×10^{-4}
<i>Xgwm604</i>	5B	(GA) ₂₉	(GA) ₂₈	5.2×10^{-4}	1.9×10^{-4} ; 1.5×10^{-3}	.	3
<i>Xgwm234</i>	5B	(GA) ₁₆ (GT) ₂₀	(GT) ₈ (GA) ₃ GG (GA) ₁₅	0	0; 5.2×10^{-4}
<i>Xgwm213</i>	5B	(GA) ₃₅	(GA) ₃₈	0	0; 5.2×10^{-4}
<i>Xgwm494</i>	6A	(GT) ₁₃	(GT) ₃ CG(GT) ₈	0	0; 5.2×10^{-4}
<i>Xgwm334</i>	6A	(GA) ₁₉	(GA) ₁₆	1.8×10^{-4}	4.3×10^{-5} ; 9.8×10^{-4}	.	1
<i>Xgwm260</i>	7A	(GA) ₂₀	(GA) ₁₁	3.7×10^{-4}	1.1×10^{-4} ; 1.3×10^{-3}	.	1	.	1	.	.	.
<i>Xgwm297</i>	7B	(GT) ₁₂ (GA) ₁₈	(GA) ₁₈ (GT) ₂₀	3.4×10^{-4}	10^{-4} ; 1.2×10^{-3}	.	.	1	1	.	.	.
<i>Xgwm537^a</i>	7B	(GT) ₁₈ (TA) ₁₃	(GT) ₂₈	7×10^{-4}	2.8×10^{-4} ; 1.8×10^{-3}	.	1	3

The last seven columns give the distribution of the mutation events at each locus, $+x$ (respectively $-x$) representing a mutation causing the gain (respectively loss) of x repeats. (—) Missing data. (·) No mutation of the corresponding size was observed. Ch, chromosome location in the variety *Chinese spring*; C.I. 95%, 95% confidence interval.

^a Loci for which the mutation rate has been estimated in a previous study (THUILLET *et al.* 2002).

sign. Among the remaining 9 loci used in the diversity analysis, no mutation was observed for 8 of them, but only the locus *Xgwm213* showed a discrepancy between its observed and predicted mutation rate from Equation 1 in the mutation-accumulation design. For locus *Xgwm126*, no direct estimation of the mutation rate was available. The predicted mutation rates at the 15 loci were significantly different among the four *T. turgidum* groups (ANOVA, $P < 0.0001$) and among loci in a given group (ANOVA, $P < 0.0001$). The estimated mutation rates at the 15 loci within each group are given in Table 2.

Estimation of θ : At equilibrium and under a SMM, both estimators $\hat{\theta}_{Hc}$ and $\hat{\theta}_F$ behaved as expected: the bias on $\hat{\theta}_{Hc}$ was removed by the correction on $\hat{\theta}_F$ (Figure 1a) and the variance of $\hat{\theta}_F$ was smaller than the variance of $\hat{\theta}_{Hc}$. At equilibrium and under a TPM, the decrease in homoplasmy resulted in an increase of H_c compared to a SMM and led to an overestimation of θ by both estima-

tors (Figure 1b). When a bottleneck occurred, its effect on the estimation of θ was less dramatic under a SMM than under a TPM, but in both cases θ was overestimated (Figure 1, c and d). For any departure from the SMM at equilibrium, a smaller bias and a smaller variance were observed for $\hat{\theta}_F$ than for $\hat{\theta}_{Hc}$. We examined the relative efficiency of both estimators, defined as the ratio of their variances. On average, $\hat{\theta}_F$ was ~ 1.5 – 2 times more efficient than $\hat{\theta}_{Hc}$. However, the efficiency of $\hat{\theta}_F$ compared to $\hat{\theta}_{Hc}$ decreased when θ increased, and the overestimation of θ by $\hat{\theta}_F$ was very important when a departure from a SMM or demographic equilibrium occurred. $\hat{\theta}_F$ was consequently chosen to estimate effective population sizes.

Departures from the SMM, demographic equilibrium, and neutrality: The locus *Xgwm213* presented a variance in allele size that was not compatible with a SMM given the number of alleles observed at this locus.

TABLE 2

Predicted mutation rates at each locus in each group ($\times 10^{-4}$)

Locus	Wild	Domesticated	Old varieties	Elite varieties
<i>Xgwm164</i>	5.9	4.4	5.3	4.7
<i>Xgwm413</i>	4.4	5.9	4.7	5
<i>Xgwm312</i>	6.8	6.8	6.5	7.4
<i>Xgwm257</i>	0.8	0.8	0.8	0.8
<i>Xgwm374</i>	4.7	3.8	2	2
<i>Xgwm2</i>	2.3	0.8	1.4	0.5
<i>Xgwm285</i>	6.2	11.3	6.5	5.6
<i>Xgwm601</i>	4.7	5	5.6	4.4
<i>Xgwm6b</i>	6.2	8	11.3	9.8
<i>Xgwm495</i>	2	2	2	1.7
<i>Xgwm126</i>	3.2	3.2	3.8	3.2
<i>Xgwm234</i>	4.4	4.7	5.3	4.1
<i>Xgwm213</i>	5.3	5.6	5.3	7.4
<i>Xgwm297</i>	5	4.1	4.7	4.7
<i>Xgwm537</i>	7.1	7.4	5.9	6.5
Mean	4.6	4.9	4.74	4.52

This departure was observed in all four groups. Loci *Xgwm312* and *Xgwm285* were also unlikely to fit a SMM hypothesis in all groups but the wild one. Locus *Xgwm6b* was not compatible with a SMM in two groups (domesticated emmer and elite durum varieties) among the four.

No evidence for a recent bottleneck was detected under the SMM hypothesis for any of the four groups (Table 3). Although a TPM assumption for all loci revealed a possible trace of a recent bottleneck in the wild group, this conclusion was dependent on the parameters chosen for the TPM and did not hold under a weaker TPM ($P = 0.1$ and $\sigma^2 = 30$), which does not provide a strong support for the existence of a recent bottleneck in this group. A deficiency in heterozygosity given the number of alleles was detected under the SMM hypothesis for both groups of durum wheat (old and elite durum varieties). Only the deficiency of heterozygosity detected in the elite durum varieties was robust to a change of model. A recent expansion of population is unlikely for this group, but this observation can be explained by a recent event of migration from a different gene pool, which could have increased the number of rare alleles without affecting the heterozygosity, thus mimicking an increase in the population size (CORNUET and LUIKART 1996). Such a scenario is indeed expected for elite durum varieties as the introduction of dwarfism genes has been realized from varieties of bread wheat (*T. t. aestivum*).

No evidence for indirect selective effect was detected using the statistic $\ln(R[\theta/\mu])$ when comparing the wild group and the domesticated emmer or when comparing old and elite durum varieties. This suggests that none of the 15 loci studied was affected by a selective sweep

either during the domestication or during more recent selection on elite durum varieties. Alternatively, all loci may have mutation rates that are high enough to erase the signature of such events. But loci *Xgwm257* and *Xgwm374* were found completely monomorphic in old durum varieties, while their heterozygosities were ~ 0.5 and 0.8 , respectively, in the domesticated emmer. They could consequently not be included in the distributions of $\ln(R[\theta/\mu])$ for the two comparisons including the group of old varieties. A selective effect during the creation of the first durum wheat varieties could have led to complete loss of polymorphism at these loci, although this observation alone is not conclusive. We note that *Xgwm257* was also monomorphic in the group of elite varieties. Among the 13 remaining loci, 1 locus (*Xgwm285*) exhibited a departure from neutrality when comparing domesticated emmer and old varieties. This effect corresponds, however, to a particularly high estimated effective population size at this locus in domesticated emmer, compared to the N_e at the other loci in this group. This locus is likely to evolve under a TPM in both groups. The detected departure from neutrality could then instead be explained by a departure from the SMM, with a greater overestimation of $\hat{\theta}_{H_e}$ in domesticated emmer than in old varieties because of a higher diversity level in domesticated emmer.

Estimation of effective population sizes: Besides loci likely to evolve under a TPM and monomorphic loci, *Xgwm601* was discarded from the estimation of N_e in the wild group because of a very high proportion of null alleles (46 individuals among 59 did not amplify). Excluding this locus and locus *Xgwm213* (TPM) from the wild group, $\hat{\theta}_F$ was highly correlated to the mutation rate in that group ($R^2 = 0.57$, $P < 0.003$, Figure 2). The same trend was observed in the three other groups ($R^2 = 0.37$ in domesticated emmer, $R^2 = 0.40$ in old durum varieties, and $R^2 = 0.44$ in elite durum varieties; $P < 0.05$). This observation supports that effective population size is rather homogenous among the remaining loci (*i.e.*, non-TPM or nonmonomorphic) and illustrates the importance of taking into account the variability of mutation rates between loci. Estimates of the mean effective population sizes per group and per locus along with an $\sim 95\%$ confidence interval are shown in Table 4. Average N_e in the wild group was $\sim 32,000$ individuals. Only $\sim 37\%$ of this effective population size has then been retained in the first domesticated form, where N_e dropped dramatically to $\sim 12,000$ individuals in the domesticated emmer. The trend was similar between domesticated emmer and old varieties, as only $\sim 50\%$ of N_e is retained from the domesticated emmer to old varieties ($N_e = 6000$). The least polymorphic group of allotetraploid wheats was the elite varieties group with a mean N_e value of 1300 corresponding to 22% of the N_e of old durum varieties and representing only 4% of the N_e of its wild ancestor. Differences in the estimated

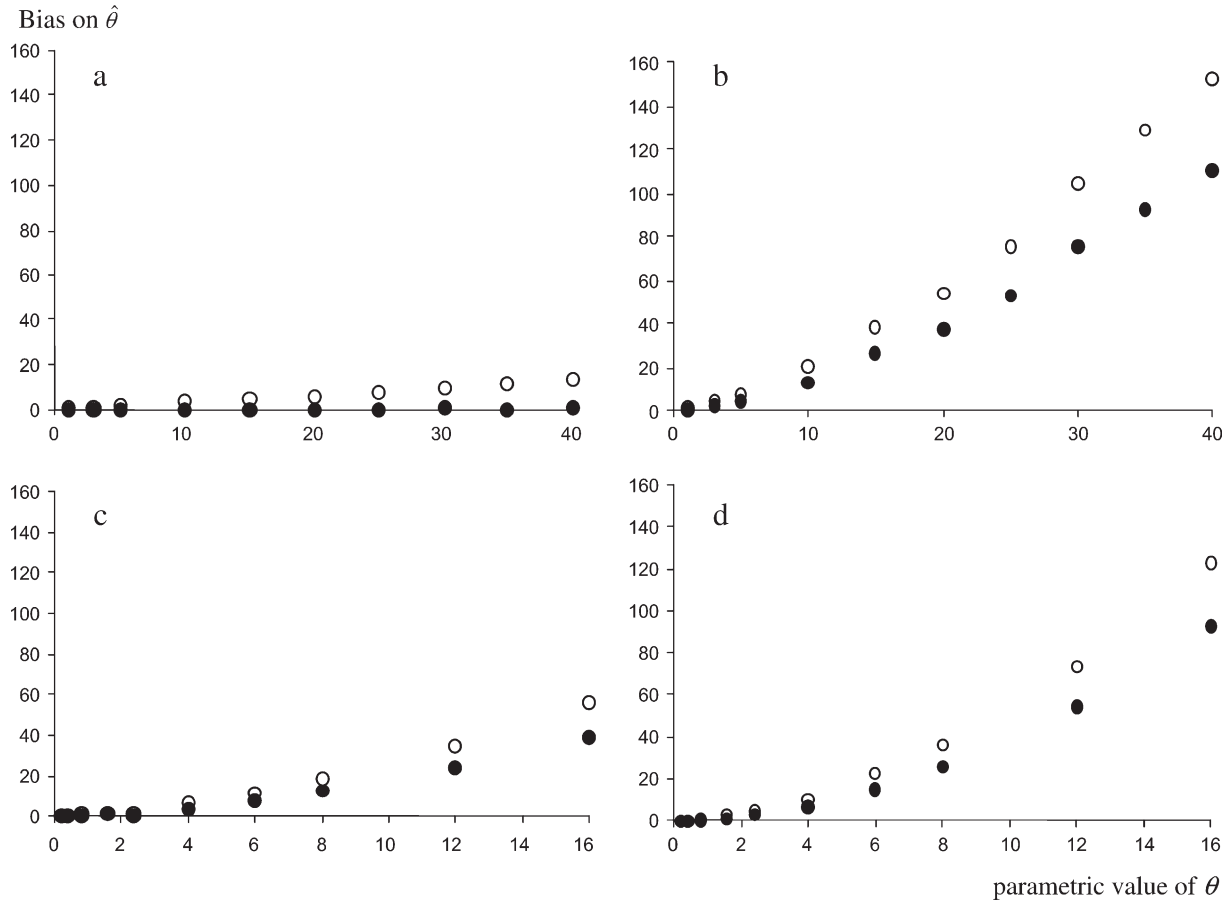


FIGURE 1.—Magnitude of the bias on $\hat{\theta}_f$ (solid circles) and on $\hat{\theta}_{fc}$ (open circles) for different values of θ . The x-axis represents the parametric value of θ used for the simulations. Each point represents the average bias estimated on the basis of 1000 simulated data sets. Simulations were realized assuming either a constant population size and (a) a SMM or (b) a TPM or a bottlenecked population and (c) a SMM or (d) a TPM.

N_e in the different groups were large and significant despite the wide confidence intervals (Table 4).

DISCUSSION

The use of microsatellites individually calibrated for their mutation rate provides a new way to obtain estimates of long-term effective population sizes. Focusing directly on the estimation of N_e allowed us to quantify the effect of the successive bottlenecks that have affected durum wheat since its domestication. We discuss below the accuracy of our method and the values we obtained in the light of the history of durum wheat.

Mutation rate estimates: Direct observation of mutation events in durum wheat is consistent with previous observations concerning microsatellite loci that typically range from 10^{-6} to 10^{-2} (WEBER and WONG 1993; SCHUG *et al.* 1997). The main conclusions were: (1) a majority of single-step mutation events affected these loci, (2) rare mutation events involving several steps could be detected, and (3) the mutation rate was length depen-

dent and allele specific rather than locus specific. The strong correlation obtained between the length of the repeated array and the mutation rate has been also observed in several other organisms (WIERDL *et al.* 1997; BRINKMANN *et al.* 1998; SCHLÖTTERER *et al.* 1998; ELLEGREN 2000a; VIGOUROUX *et al.* 2002).

As the allele length appeared to be a crucial factor for the variation of the mutation rate in our set of loci, we used this property to predict mutation rates in the different populations. This can be inaccurate if other factors have an impact on the mutation rate as well (*i.e.*, repeated motif, chromosomal position; THUILLET *et al.* 2004). The size of the repeated motif and the structure of the alleles (perfect or imperfect) have both been shown to have a substantial effect on the mutation rate of microsatellites (CHAKRABORTY *et al.* 1997; BRINKMANN *et al.* 1998). In our case, we used only dinucleotide loci. Although perfect microsatellites yielded a stronger correlation between mutation rate and length of the repeated array than when both perfect and imperfect loci were included, we kept the two imperfect loci to

TABLE 3

Test for mutation-drift equilibrium using the method of CORNUET and LUIKART (1996)

	SMM		TPM	
	<i>E/D</i>	<i>P</i> -value	<i>E/D</i>	<i>P</i> -value
Wild	9/4	0.59	10/3	0.04
Domesticated	4/7	0.91	6/5	0.90
Old varieties	3/7	0.03	4/6	0.85
Elite varieties	0/10	<0.001	0/10	<0.001

E/D, number of loci showing an excess/deficiency heterozygosity given the observed number of alleles. The associated *P*-value for a Wilcoxon signed-rank test (two tails) is given.

build Equation 1 since this did not prevent observing a correlation between $\hat{\theta}_F$ and predicted mutation rate. As 83% of the variation in μ in the accumulation design was explained by the number of repeats and the equation relied on loci also used in the diversity survey, we believe that our prediction for μ is likely to be fairly accurate. Miscellaneous and unidentified factors are possible, however, as suggested by the existence of three outlier loci. Two of them were not used for the diversity analysis, and the last one, *Xgwm213*, was discarded from the estimation of N_e in all groups since it was suspected to mutate under a TPM. Another source of error in the prediction of the mutation rate is a bad estimation of the number of repeats of the alleles. For example, in maize some microsatellites have been shown to present a continuous variation of their size due to a high proportion of insertions or deletions (MATSUOKA *et al.* 2002). This kind of situation would make it difficult to deduce the number of repeats from the size of the alleles. In our case, the structure of the loci was considered to be conserved between individuals as it appeared to be mostly conserved between the genomes of the bread wheat genotype *Chinese spring* and the durum wheat genotype *Sham I*. Furthermore, the allelic distributions presented either even or odd sizes but not both at the same locus, suggesting an evolution by steps rather than by insertions/deletions.

Effective population size estimates: Significant differences existed in the predicted mutation rates between groups and between loci. Therefore, the estimation of effective population sizes was realized using predicted mutation rates per locus and for each group. Apart from the error due to the prediction of μ , here assumed to be relatively low, an important source of error in the estimation of N_e comes from the estimation of the parameter θ . The corrected estimator $\hat{\theta}_F$ appeared to be more efficient than $\hat{\theta}_{He}$ even when a departure from the SMM and demographic equilibrium assumption occurs and was consequently chosen to estimate θ in our diversity study. The use of $\hat{\theta}_F$, however, can be extremely misleading when such a departure from the assumptions

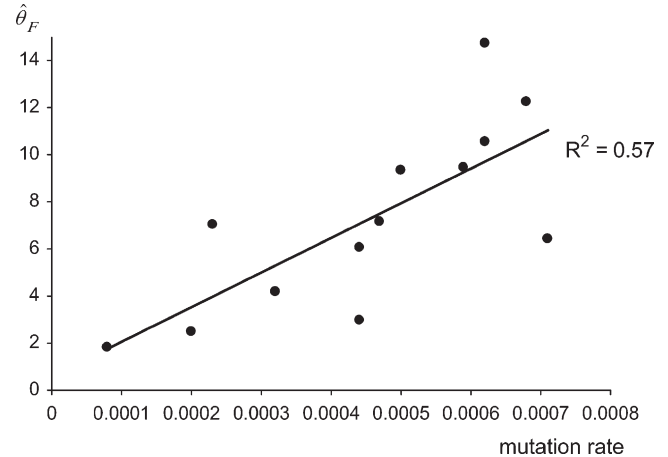


FIGURE 2.—Correlation between $\hat{\theta}_F$ and the predicted mutation rate in the wild group of wheat ($R^2 = 0.56$, $P < 0.003$).

occurs (especially when the diversity level is high). In that case, our values of N_e would be overestimated.

We detected four loci that are likely to follow a TPM in at least one group. The tests used to discriminate between SMM and TPM were realized under a demographic equilibrium hypothesis. If the tested populations have actually experienced a recent bottleneck, we might have missed loci for which the TPM hypothesis is likely (THUILLET *et al.* 2004). Our conclusion concerning the four TPM loci would still hold in that case. These loci exhibited a particularly high value of $\hat{\theta}_F$ (data not shown), supporting the existence of multistep mutation events during their evolution. Note that multisteps have been observed in the mutation-accumulation design for the locus *Xgwm6b*. These four loci were excluded from the N_e estimation in groups where the SMM was unlikely. The remaining loci were assumed to have evolved under a SMM. This does not exclude the existence of a minor departure from the SMM, but it should also have a less dramatic effect on the estimation of θ .

The existence of a bottleneck in the studied populations is also likely to lead to an overestimation of θ by $\hat{\theta}_F$. No strong evidence for the existence of a bottleneck in any group was detected when using Bottleneck software (CORNUET and LUIKART 1996). However, the deficiency in Nei's heterozygosity given the number of alleles in the elite varieties of durum wheat was highly significant and robust to the assumed pattern of evolution (SMM or TPM). The proposed scenario of a recent introduction of alleles from bread wheat is plausible and in that case we probably underestimated θ compared to what it would be if the population was at equilibrium state. The magnitude of this effect is difficult to measure, but it should be noted that the number of varieties of bread wheat that are known to have been used to introduce dwarfism genes is small. Thus, the impact of this introduction may be small.

Demographic bottlenecks in durum wheat: A continu-

TABLE 4
Estimates of effective population sizes (N_e)

Locus	Wild	Domesticated	Old varieties	Elite varieties
<i>Xgwm164</i>	38272	2557	679	748
<i>Xgwm413</i>	5022	4208	1328	943
<i>Xgwm312</i>	55155	22735	1684	244
<i>Xgwm257</i>	10236	2945	0	0
<i>Xgwm374</i>	27159	6469	0	13
<i>Xgwm2</i>	53596	10005	213	51
<i>Xgwm285</i>	45268	69613	832	267
<i>Xgwm601</i>	<i>249</i>	16171	1357	532
<i>Xgwm6b</i>	87630	27839	9391	7241
<i>Xgwm495</i>	7653	23797	18216	2880
<i>Xgwm126</i>	13900	1311	3510	2360
<i>Xgwm234</i>	20871	2410	11730	3280
<i>Xgwm213</i>	120565	13391	844	1992
<i>Xgwm297</i>	43536	36041	7256	1086
<i>Xgwm537</i>	14579	26456	5973	1367
Mean	32529	12034	5965	1326
95% C.I.	[20739; 45991]	[5790; 19300]	[2831; 9556]	[689; 2031]

Underlined values were not included in the calculation of the mean effective population size of a group. Values in boldface type represent loci detected as evolving in a TPM manner. Values in italics represent loci displaying a high frequency of null alleles, and when a locus is fixed in a group, the value zero is given to N_e .

ous decrease of effective population sizes was observed from the wild form of wheat to the current cultivated subspecies of allotetraploid wheat. The effective population size of each group is the harmonic mean of the effective population sizes realized throughout the history of each subspecies. A subspecies with a smaller effective population size is considered to have experienced a stronger or a more intense bottleneck. *T. t. dicoccoides* is considered as the current lineage originating from the wild ancestor of AB allotetraploid wheats. The fact that we detected no strong evidence for the existence of a past reduction of population size in this group does not mean that the effective population size has remained constant for 12,000 years. If a strong reduction of size occurred a long time ago and demographic size subsequently stayed constant, this group might have recovered an equilibrium state and could exhibit no footprint of bottleneck. This could have happened if a reduction of the natural geographic area of the wild subspecies has occurred, due, for instance, to an expansion of cultivated surfaces. In that case, it is possible that the effective size estimated in *T. t. dicoccoides* represents a minimum estimate of the diversity in the wild ancestor at the time of domestication.

The current effective size of *T. t. dicoccum*, which was the first domesticated wheat derived from the wild form, was only 37% of that of its ancestor *T. t. dicoccoides*. In terms of Nei's heterozygosity, 95% of the diversity present in *T. t. dicoccoides* was retained in *T. t. dicoccum* (Figure 3). Compared to other species, this value was very high as ~60–70% of genetic diversity was retained through domestication in maize, sorghum, rice, oats,

or pearl millet (BUCKLER *et al.* 2001). This discrepancy between the decrease in H_e and the decrease in N_e may be due to the high mutation rates of microsatellites that can recover diversity much quicker than gene nucleotide sequences, on which these other studies were based. This is also true for subsequent reductions and shows that the bottleneck effect may be largely underestimated when one focuses only on summary statistics of diversity especially when (new) mutation accounts for a substantial portion of the observed polymorphism (Figure 3). After the domestication event, *T. t. dicoccum* spread out of the Fertile Crescent and was then replaced by *T. t.*

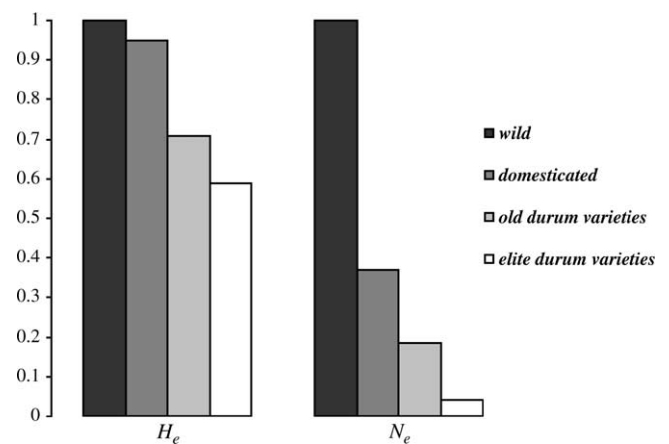


FIGURE 3.—Reductions of heterozygosity (H_e) and effective population size (N_e) from the wild form, *T. t. dicoccoides*, to the elite varieties of durum wheat. Values are shown relative to the value in the wild form that is retained in the subsequent population.

durum populations (NESBITT and SAMUEL 1996; ZOHARY and HOPF 2000). So, most of the populations of *T. t. dicoccum* that were present after the domestication have disappeared. The current diversity and N_e values of this group are then more difficult to interpret as they likely result from the succession of several demographic events. However, the current value of N_e in this group is likely to be smaller than what it was after the domestication. Therefore the reduction of N_e imposed by the free-threshing acquisition could be underestimated.

On the other hand, N_e in old durum varieties could also be underestimated since a number of landraces might have disappeared. Elite varieties of our study come mostly from the French variety catalog but are expected to sample adequately the diversity in modern worldwide material due to a low number of founders of short stature at the international level and subsequent intense exchange between countries (P. GRIGNAC, personal communication). The reduction of the effective population size in that group (-78% relative to old varieties) was dramatic. This value is representative of the bottleneck associated with the selection for semi-dwarf stature and to intensive cultivation practices.

Footprints of selection in durum wheat? None of the polymorphic loci showed obvious departure from neutrality. However, two loci (*Xgwm257* and *Xgwm374*) were completely monomorphic in old varieties of durum wheat. Both are located on the short arm of chromosome 2B. The gene *Tg2* (tenacious glumes), also located on the same chromosomal arm, is a candidate gene for a QTL involved in the acquisition of free threshing during the transition from *T. t. dicoccum* to *T. t. durum* (SIMONETTI *et al.* 1999).

A first explanation for the absence of polymorphism at loci *Xgwm257* and *Xgwm314* could be that these loci experienced a reduction of diversity that is comparable to the rest of the genome (through a demographic bottleneck) but exhibit a low mutation rate and therefore did not recover any diversity. Under this assumption, estimates of their mutation rate should be low as well as their level of diversity in *T. t. dicoccoides*. Locus *Xgwm257* presented indeed a low heterozygosity in the wild group (0.69, whereas the average for the group is 0.85) and had a low predicted mutation rate (0.8×10^{-4}) with no mutation detected in the mutation-accumulation design. Locus *Xgwm374*, on the other hand, had a high mutation rate (2×10^{-4} – 4.7×10^{-4}) and high heterozygosity ($H_e = 0.91$) in *T. t. dicoccoides*. Thus a mere demographic effect appears unlikely to explain patterns of microsatellite variation observed at this locus.

The marker *Xrsq805*, located on the distal region of chromosome 2B, is reported by SIMONETTI *et al.* (1999) as strongly associated with threshability in tetraploid wheat and as marking a QTL corresponding to locus *Tg2*. By comparison of different existing maps of wheat (RÖDER *et al.* 1998a; SIMONETTI *et al.* 1999; PAILLARD *et al.* 2003), we inferred the genetic position of *Xgwm374* and *Xgwm257* loci relative to *Xrsq805*. *Xgwm257* should be ~ 37 cM more proximal than the putative position of *Tg2*, and *Xgwm374* was itself located 35 cM more proximal than *Xgwm257* (72 cM from *Xrsq805*). By comparison with physical maps of wheat, it appears that *Xgwm257* and *Tg2* belong to the same physical bin (between 84 and 100% of the distal part of the short arm; SOURDILLE *et al.* 2004), whereas *Xgwm374* belongs to the adjacent proximal physical bin. The genetic distance between both microsatellite markers covers $\sim 40\%$ of the estimated physical length of the chromosomal arm (RÖDER *et al.* 1998b; SOURDILLE *et al.* 2004). Consequently, the marker *Xgwm374* is actually located relatively far from the *Tg2* gene. A direct effect of the gene *Tg2* on this marker would require a strong selective effect spanning a very large region of the genome. On the other hand, if *Tg2* is not involved in the selection process that seems necessary to explain the loss of polymorphism at *Xgwm374*, another candidate gene could be located closer to the microsatellite marker. It would then be interesting to investigate patterns of polymorphism in the surrounding region of this marker.

No trace of selection attributable directly to domestication was detected in our data set. We also did not detect genomic regions with a marked decrease of effective sizes due to selection during key stages of modern breeding. The main explanation is probably the low number of markers used in this study. Even with strong linkage drag, as expected in a selfing species, a greater number of markers are required to cover the 14 chromosomes of allotetraploid wheats.

We thank M.-H. Muller for kindly providing the coalescence simulation program used in this article. We thank R. M. Clark and two anonymous reviewers for helpful comments on a previous version of this article. This work was supported by grants from the "Bureau des Ressources Génétiques" and the "Comité Technique Permanent de La Sélection des Plantes Cultivées."

LITERATURE CITED

- BRINKMANN, B., M. KLINTSCHAR, F. NEUHUBER, J. HUHNE and B. ROLF, 1998 Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62**: 1408–1415.
- BRODEHE, J., A. P. MØLLER and H. ELLEGREN, 2004 Individual variation in microsatellite mutation rate in barn swallows. *Mutat. Res.* **545**: 73–80.
- BUCKLER, IV, E. S., J. M. THORNSBERRY and S. KRESOVICH, 2001 Molecular diversity, structure and domestication of grasses. *Genet. Res.* **77**: 213–218.
- CASELLA, G., and R. L. BERGER, 1990 *Statistical Inference*. Duxbury Press, Belmont, CA.
- CHAKRABORTY, R., M. KIMMEL, D. N. STIVERS, L. J. DAVISON and R. DEKA, 1997 Relative mutation rates at di-, tri-, and tetranucleotides microsatellite loci. *Proc. Natl. Acad. Sci. USA* **94**: 1041–1046.
- CORNUET, J. M., and G. LUIKART, 1996 Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**: 2001–2014.
- DI RIENZO, A., A. C. PETERSON, J. C. GARZA, A. M. VALDES, M. SLATKIN *et al.*, 1994 Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl. Acad. Sci. USA* **91**: 3166–3170.

- ELLEGREN, H., 2000a Heterogeneous mutation processes in human microsatellite DNA sequences. *Nat. Genet.* **24**: 400–402.
- ELLEGREN, H., 2000b Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* **16**: 551–558.
- HUDSON, R. R., 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **9**: 1–44.
- KAUER, M. O., D. DIERINGER and C. SCHLÖTTERER, 2003 A microsatellite variability screen for positive selection associated with the “out of Africa” habitat expansion of *Drosophila melanogaster*. *Genetics* **165**: 1137–1148.
- KIMMEL, M., and R. CHAKRABORTY, 1996 Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theor. Popul. Biol.* **50**: 345–367.
- LEHMANN, T., W. A. HAWLEY, H. GREBERT and F. H. COLLINS, 1998 The effective population size of *Anopheles gambiae* in Kenya: implications for population structure. *Mol. Biol. Evol.* **15**: 264–276.
- MAIER, U., 1996 Morphological studies of free threshing wheat ears from a Neolithic site in southwest Germany, and the history of naked wheats. *Veg. Hist. Archaeobot.* **5**: 39–55.
- MATSUOKA, Y., S. E. MITCHELL, K. KRESOVICH, M. GOODMAN and J. DOEBLEY, 2002 Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. *Theor. Appl. Genet.* **104**: 436–450.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- NEI, M., and D. GRAUR, 1984 Extent of protein polymorphism and the neutral mutation theory. *Evol. Biol.* **17**: 73–118.
- NEI, M., T. MARUYAMA and R. CHAKRABORTY, 1975 The bottleneck effect and genetic variability in populations. *Evolution* **29**: 1–10.
- NESBITT, M., and D. SAMUEL, 1996 From staple crop to extinction? The archeology and history of the hulled wheat, pp. 41–100 in *First International Workshop on Hulled Wheats*, edited by S. PADULOSI, K. HAMMER and J. HELLER. IPGRI, Castelvecchio Pascoli, Tuscany, Italy.
- OHTA, T., and M. KIMURA, 1973 A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* **22**: 201–204.
- PAILLARD, S., T. SCHNURBUSCH, M. WINZELER, M. MESSMER, P. SOURDILLE *et al.*, 2003 An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **107**: 1235–1242.
- PASCUAL, M., C. F. AQUADRO, V. SOTO and L. SERRA, 2001 Microsatellite variation in colonizing and paleartic populations of *Drosophila subobscura*. *Mol. Biol. Evol.* **18**: 731–740.
- RÖDER, M. S., K. KORZUN, J. P. K. WENDEHAKE, M.-H. TIXIER, P. LEROY *et al.*, 1998a A microsatellite map of wheat. *Genetics* **149**: 2007–2023.
- RÖDER, M. S., V. KORZUN, B. S. GILL and M. W. GANAL, 1998b The physical mapping of microsatellite markers in wheat. *Genome* **41**: 278–283.
- SAS, 1999–2000 *SAS/STAT User's Guide*, Version 8.1. SAS Institute, Cary, NC.
- SCHLÖTTERER, C., 2000 Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**: 365–371.
- SCHLÖTTERER, C., 2002 A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* **160**: 753–763.
- SCHLÖTTERER, C., R. RITTER, B. HARR and G. BREM, 1998 High mutation rate of a long microsatellite allele in *Drosophila melanogaster* provides evidence for allele-specific mutation rates. *Mol. Biol. Evol.* **15**: 1269–1274.
- SCHUG, M. D., T. F. C. MACKAY and C. F. AQUADRO, 1997 Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nat. Genet.* **15**: 99–102.
- SIMONETTI, M. C., M. P. BELLOMO, G. LAGHETTI, P. PERRINO, R. SIMEONE *et al.*, 1999 Quantitative trait loci influencing free-threshing habit in tetraploid wheats. *Genet. Resour. Crop Evol.* **46**: 267–271.
- SOURDILLE, P., S. SINGH, T. CADALEN, G. L. BROWN-GUEDIRA, G. GAY *et al.*, 2004 Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* **4**: 12–25.
- SPONG, G., M. JOHANSSON and M. BJÖRKLUND, 2000 High genetic variation in leopards indicates large and long-term stable effective population size. *Mol. Ecol.* **9**: 1773–1782.
- TAI, T. H., and S. D. TANKSLEY, 1990 A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant Mol. Biol. Rep.* **8**: 297–303.
- THUILLET, A. C., D. BRU, J. DAVID, P. ROUMET, S. SANTONI *et al.*, 2002 Direct estimation of mutation rate for 10 microsatellite loci in durum wheat, *Triticum turgidum* (L.) Thell. ssp. *durum* Desf. *Mol. Biol. Evol.* **19**: 122–125.
- THUILLET, A. C., T. BATAILLON, P. SOURDILLE and J. L. DAVID, 2004 Factors affecting polymorphism at microsatellite loci in bread wheat [*Triticum aestivum* (L.) Thell.]: effects of mutation processes and physical distance from the centromere. *Theor. Appl. Genet.* **108**: 368–377.
- VIGOUROUX, Y., J. S. JAQUETH, Y. MATSUOKA, O. S. SMITH, W. D. BEAVIS *et al.*, 2002 Rate and pattern of mutation at microsatellite loci in maize. *Mol. Biol. Evol.* **19**: 1251–1260.
- WANG, R. L., A. STEC, J. HEY, L. LUKENS and J. DOEBLEY, 1999 The limits of selection during maize domestication. *Nature* **398**: 236–239.
- WEBER, J. L., and C. WONG, 1993 Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**: 1123–1128.
- WIERDL, M., M. DOMINSKA and T. D. PETES, 1997 Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**: 769–779.
- XU, H., and Y. X. FU, 2004 Estimating effective population size or mutation rate with microsatellites. *Genetics* **166**: 555–563.
- XU, X., M. PENG and Z. FANG, 2000 The direction of microsatellite mutations is dependent upon allele length. *Nat. Genet.* **24**: 396–399.
- ZOHARY, D., and M. HOPF, 2000 *Domestication of Plants in the Old World*. Oxford University Press, New York.
- ZOUROS, E., 1979 Mutation rates, population sizes and amounts of electrophoretic variation of enzyme loci in natural populations. *Genetics* **92**: 623–646.

Communicating editor: S. R. McCouch

