Sequence Polymorphism in Polyploid Wheat and Their D-Genome Diploid Ancestor

Katherine S. Caldwell,* Jan Dvorak,† Evans S. Lagudah,‡ Eduard Akhunov,‡
Ming-Cheng Luo,† Petra Wolters* and Wayne Powell*,[1]

*E. I. DuPont Nemours, DuPont Agricultural Products, Newark, Delaware 19711, †Department of Agronomy and Range Science,
University of California, Davis, California 95616 and ‡Commonwealth Scientific and Industrial Research Organization,
Division of Plant Industry, Canberra, Australian Territory, 2601 Australia

ABSTRACT

Sequencing was used to investigate the origin of the D genome of the allopolyploid species *Triticum aestivum* and *Aegilops cylindrica*. A 247-bp region of the wheat D-genome *Xwye838* locus, encoding ADP-glucopyrophosphorylase, and a 326-bp region of the wheat D-genome *Gss* locus, encoding granule-bound starch synthase, were sequenced in a total 564 lines of hexaploid wheat (*T. aestivum*, genome AABBDD) involving all its subspecies and 203 lines of *Aegilops tauschii*, the diploid source of the wheat D genome. In *Ae. tauschii*, two SNP variants were detected at the *Xwye838* locus and 11 haplotypes at the *Gss* locus. Two haplotypes with contrasting frequencies were found at each locus in wheat. Both wheat *Xwye838* variants, but only one of the *Gss* haplotypes seen in wheat, were found among the *Ae. tauschii* lines. The other wheat *Gss* haplotype was not found in either *Ae. tauschii* or 70 lines of tetraploid *Ae. cylindrica* (genomes CCDD), which is known to hybridize with wheat. It is concluded that both *T. aestivum* and *Ae. cylindrica* originated recurrently, with at least two genetically distinct progenitors contributing to the formation of the D genome in both species.

P OLYPLOIDY, the presence of more than one genome per cell, is a common speciation mechanism in the plant kingdom and a large proportion of plants are either recent or ancient polyploids (STEBBINS 1950; SOLTIS and SOLTIS 1993; LEITCH and BENNETT 1997; OTTO and WHITTON 2000; WENDEL 2000). Numerous studies have also demonstrated that polyploidy has been significant in the evolution of many vertebrates and other eukaryotes (WOLFE and SHIELDS 1997; PEBUSQUE *et al.* 1998). Polyploids are often classified into allopolyploids, resulting from interspecific hybridization of two fully differentiated genomes, or autopolyploids, arising from intraspecific chromosome doubling (STEBBINS 1947). Both modes of polyploidization result in duplication of genetic loci and, in allopolyploids, in the creation of homeologous loci contributed by different donor taxa at the time of polyploid formation. It has often been asked why polyploids are so common and successful (SOLTIS and SOLTIS 1999). LEITCH and BENNETT (1997) proposed that the success of polyploids is associated with their mode of formation and the extent of genetic divergence between the parents. Thus, polyploid species formed recurrently from genetically divergent parents capture genomic diversity from multiple progenitor populations, thereby increasing genetic diversity over that achieved via a single hybridization event (SOLTIS and SOLTIS 1993, 1999). Recurrent or multiple polyploidization events therefore may be an important source of novel genetic diversity that allows polyploids to adapt to new ecological niches or enhance their agricultural performance (LEVIN 1980; OSBORN *et al* 2003).

Allopolyploidy is prominent in plants, particularly in agriculturally important crops (SALAMINI *et al.* 2002). One of these, bread wheat (*Triticum aestivum*), originated by hybridization of cultivated allotetraploid emmer wheat (*T. turgidum* ssp. *dicoccum*, $2n = 4x = 28$, genomes *AABB*) with diploid *Aegilops tauschii* ($2n = 2x = 14$, genome *DD*; KIHARA 1944; MCFADDEN and SEARS 1946). The latter has two subspecies: ssp. *strangulata* and ssp. *tauschii* (VAN SLAGERN 1994). For several isozymes, nearly fixed polymorphisms were detected between the two subspecies and, in each case, only the *Ae. tauschii* ssp. *strangulata* allele was found in wheat (NISHIKAWA 1973; JAASKA 1978, 1981; NISHIKAWA *et al.* 1980). *Aegilops cylindrica* originated by hybridization of *Aegilops caudata* (genome *CC*) with *Ae. tauschii* (KIHARA 1931) and is known to hybridize spontaneously with wheat (GANDILYAN and JAASKA 1980; SNYDER *et al.* 2000). The chromosomes of the D genome of *Ae. cylindrica* and those of the D genome of *T. aestivum* are homologous and pair at metaphase I in artificially produced *T. aestivum* × *Ae. cylindrica* hybrids (KIHARA 1944). Naturally formed

[1]*Corresponding author:* Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK.   E-mail: wpowel@scri.sari.ac.uk

hybrids between *Ae. cylindrica* and wheat occasionally produce seeds, presumably via spontaneous cross-pollination, which could facilitate gene flow between the D genomes of the two species (GANDILYAN and JAASKA 1980; SEEFELDT *et al.* 1998; GUADAGNUOLO *et al.* 2001).

Indirect measures of sequence polymorphism based on restriction fragment length polymorphism (RFLP) of single-gene loci (DVORAK *et al.* 1988) and restriction site variation of low-copy DNA (TALBERT *et al.* 1998) have been used as evidence for the recurrent origin of wheat polyploids and the existence of gene flow from parental species to polyploids. TALBERT *et al.* (1998) concluded that hexaploid wheat was formed at least twice but further information on the exact origin of the D genome is still outstanding. Such information can be derived from direct analysis of DNA sequence polymorphism and from the identification of haplotypes representing several physically linked single nucleotide polymorphisms (SNPs; HEY 1998). Such approaches are now being used to study sequence divergence between sibling species (HILTON and GAUT 1998); to investigate genetic bottlenecks leading to domestication of crop plants, particularly maize (EYRE WALKER *et al.* 1998); and to quantify biological factors influencing the patterns of genetic diversity observed in various organisms (CUMMINGS and CLEGG 1998; FILATOV and CHARLESWORTH 1999; LIN *et al.* 2001).

The polyploid origin of wheat from three ancestral genomes, together with the associated triplication of genetic information, makes SNP and haplotype studies in polyploid species more challenging than comparable studies in diploids. Analysis of hexaploid wheat requires the development of genome-specific primers to ensure that only one of the three genomes (A, B, and D) is amplified to reveal the haplotypes present in the specific genome being studied. Studies focusing on the organization of sequence polymorphism into haplotypes in plants are in their infancy (WHITE and DOEBLEY 1999; TENAILLON *et al.* 2001; RAFALSKI 2002). However, since sequence variation reflects the shared population history of contiguous DNA segments, it provides a powerful approach to unravel the evolutionary history of crop plants. Such an approach, coupled with a comprehensive sampling strategy, is well suited to determine the number of independent parental lineages contributing to the formation of a polyploid species. We report here a study of D-genome-specific haplotype data at genes encoding ADP-glucose pyrophosphorylase (*Xwye838*; AINSWORTH *et al.* 1995) and granule-bound starch synthase (*Gss*) to investigate and quantify the recurrent origin of *T. aestivum* and *Ae. cylindrica* on the basis of extensive sampling of hexaploid, tetraploid, and diploid wheat. We conclude that both *T. aestivum* and *Ae. cylindrica* originated recurrently with at least two genetically distinct progenitors contributing to the formation of the D genome in both species.

MATERIALS AND METHODS

Plant material: A total of 564 hexaploid wheat lines, 203 *Ae. tauschii* lines, and 70 lines of *Ae. cylindrica* were studied. The geographical origins of the lines are listed in the supplemental data at http://www.genetics.org/supplemental/.

DNA extraction: Nuclear DNA of most of the wheat and *Ae. tauschii* lines and all *Ae. cylindrica* lines was isolated according to a procedure described by DVORAK *et al.* (1988). The remaining genomic DNA was isolated as follows. Leaf tissue was ground to a fine powder with a chilled mortar and pestle in the presence of glass beads and liquid nitrogen. In a 15-ml polypropylene conical tube, 3 ml of Plant DNAzol (Life Technologies) for every gram of fresh tissue was added and mixed by shaking. An equal volume of phenol-chloroform-isoamyl was added, vortexed briefly, and centrifuged for 7 min at $3000 \times g$. The upper phase was transferred to a new 15-ml conical polypropylene tube. DNA was precipitated with 0.7 vol cold isopropanol, removed with a Pasteur pipette hook, and placed in a clean Eppendorf tube with 1 ml 70% ethanol. The sample was centrifuged for 10 min at $3000 \times g$ and the ethanol was decanted. The pellet was dried of all residual ethanol and resuspended in 300 µl TE with 2 µl RNAse (10 mg/ml).

PCR amplification: The *Xwye838* and *Gss* loci were amplified from the hexaploid variety Chinese Spring, cloned into Pgem-T, and transformed into DH10b electrocompetent (Life Technologies) cells. The sense-strand primer 5′-GTACGGGCTAG TGAAGTTCG-3′ and antisense-strand primer 5′-GGGAGTTT GTTTGCTCGC-3′ were designed directly from the GenBank sequence and used for the amplification of *Xwye838* from genomic DNA. *Gss* was amplified by the Sun1 primers described by SHARIFLOU and SHARP (1999). A sample of 13–15 clones per gene was sequenced to determine nucleotide diversity at homeologous loci. Genome-specific primers were designed utilizing substitutions and indels and were tested for specificity to the three genomes of wheat using nullisomic/tetrasomic substitution lines. PCR amplifications were performed in a 25-µl volume containing 1× PE buffer, 1.5 mM MgCl$_2$, 0.2 mM of each dNTP, 0.2 M of both forward and reverse primers, 0.65 units AmpliTaq Gold, and 50 ng genomic DNA. PCR reactions with the following cycling conditions were performed in a Perkin-Elmer (Norwalk, CT) 9700 thermocycler: 96° for 10 min, 40 cycles of 96° for 1 min, 55° for 1 min, and 72° for 2 min, concluded with a final extension of 72° for 7 min. The resulting amplicons were separated on 1% agarose gel 1× TBE to determine quality and quantity of product purified with USB ExoSAP-It.

DNA sequencing: Left primer 5′-TGGTTAGCACTGGTCT TACC and right primer 5′-CAACGGCATCGTCACGAGAGC were used to PCR amplify a portion of the *Xwye838-1D* locus and left primer 5′-CAATGTTTAGTCAGGGCAGC-3′ and right primer 5′-AAGTGGAAACGAGATTAAGC-3′ were used to PCR amplify a portion of the *Gss* locus specifically from the D genome. The addition of T3 (5′-AATTAACCCTCACTAAAGGG-3′) and T7 (5′-GTAATACGACTCACTATAGGGC-3′) tags to the forward and reverse genome-specific primers, respectively, for PCR amplification enabled direct sequencing from the PCR amplicons with T3 and T7 primers. ABI PRISM dye terminator cycle sequencing ready reaction kit with AmpliTaq FS DNA polymerase (PE Applied Biosystems, Foster City, CA) was used for the sequencing reactions. The sequences were analyzed on an ABI 377 automated sequencer (PE Applied Biosystems) and aligned with Sequencher (Gene Codes, Ann Arbor, MI).

Data analysis: Two measures of nucleotide polymorphism were calculated (TAJIMA 1983): nucleotide diversity (π) based on the average number of pairwise comparisons in a sample (NEI 1987) and Watterson's estimator (θ) based on the number of segregating sites (WATTERSON 1975). Evidence of past

	23	34	38	41	42	46	53	54	57	66	72	81	82	87	93	96	101	115	118	120	126	132	145	178	179	181	182	183	184
(*)							*	*		*								*	*				*						
Haplotype 1	A	C	C	A	T	G	G	G	A	C	G	T	C	A	T	C	G	G	A	C	G	T	A	G	A	A	A	G	G
Haplotype 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 4	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	T	.	.	.	.	.	.	.	.	.
Haplotype 6	.	T	T	C	:	.	C	.	.	T	T	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	.	.	:
Haplotype 7	.	.	.	.	.	A	.	A	T	.	.	G	.	G	G	G	.	A	G	.	.	C	.	A	G	G	.	C	:
Haplotype 8	.	.	.	.	.	A	.	A	T	.	.	G	.	G	G	G	.	A	G	.	.	C	.	A	G	G	.	C	:
Haplotype 9	.	.	.	.	.	A	T	.	T	.	A	.	.	.	G	.	.	A	G	.	.	C	.	A	G	G	G	C	:
Haplotype 10	.	.	.	.	.	A	.	A	T	.	.	G	.	G	G	G	.	A	G	.	.	C	C	.	.	G	.	.	:
Haplotype 11	.	.	.	.	.	A	.	A	T	.	.	G	.	G	G	G	.	A	G	.	.	C	C	.	.	G	.	.	:
Haplotype 12	G	.	.	.	.	A	.	A	T	.	.	G	.	G	G	G	.	A	G	.	.	C	C	.	.	G	.	.	:

	185	186	187	188	189	191	194	196	197	198	199	200	201	202	203	204	205	208	227	235	236	257	271	278	279	292	300	301
(*)																						*						
Haplotype 1	A	A	G	T	T	T	T	C	A	T	G	G	A	G	C	G	T	A	C	:	:	C	T	:	:	T	C	:
Haplotype 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.
Haplotype 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	:	:	.	.	:	:	.	.	:
Haplotype 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	:	:	.	.	:	:	.	.	:
Haplotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	:	:	.	.	:	:	.	.	:
Haplotype 6	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	A	.	.	T	.	G	T	A	T	T
Haplotype 7	:	:	:	:	.	C	.	.	:	:	:	:	:	:	:	:	:	:	:	.	.	.	T	.	.	.	A	.
Haplotype 8	:	:	:	:	.	C	.	.	:	:	:	:	:	:	:	:	:	:	:	.	.	.	T	.	.	.	.	.
Haplotype 9	:	:	:	:	.	C	.	.	:	:	:	:	:	:	:	.	C	.	A	.	A	T	.	.	.	A	.	T
Haplotype 10	:	:	:	:	.	C	G	A	:	:	:	:	:	:	:	:	:	:	:	.	.	.	.	.	.	.	.	.
Haplotype 11	:	:	:	:	.	C	G	A	:	:	:	:	:	:	:	:	:	:	:	.	.	.	.	C	.	.	.	.
Haplotype 12	:	:	:	:	.	C	G	A	:	:	:	:	:	:	:	:	:	:	:	.	.	.	.	.	.	.	.	T

FIGURE 1.—Amplification of a 326-bp region of *Gss* with D-genome-specfic primers identified 12 haplotypes. Only the polymorphic consensus positions are indicated. (*) Substitutions that were nonsynonymous. (:) Missing bases at regions of insertion/deletion events.

selection was investigated using Tajima's *D* test (TAJIMA 1989). Estimates of polymorphism and tests of selection were calculated with dnaSP version 3.53 (ROZAS and ROZAS 1999).

RESULTS

Nucleotide diversity at homeologous loci: *Xwye838* and *Gss* sequences amplified from hexaploid Chinese Spring and representing different wheat genomes were compared. For each gene, two clearly defined haplotypes that correspond to the A and D genomes were identified (see below). The high number of differences observed between the homeologous loci suggests that the missing haplotype representing the B genome (based on nullisomic/tetrasomic analysis) could be a result of a lack of primer specificity. In the case of *Xwye838* a total of 50 substitutions were identified, with an average of 1 substitution every 19 bp. A similar frequency of substitution was observed in the *Gss* sequence (1/20 bp); however, the number of insertion/deletions (indels) differed between the two sequences (7 for *Xwye-838* and 10 for *Gss*). On average, we detected 1 substitution/20 bp and one indel/84 bp of sequence between homeologs. Estimates of divergence per synonymous nucleotide site, expressed as K_s values, are 0.0623 for *Gss* and 0.0971 for *Xwye838* (no nonsynonymous substitutions were detected). This divergence between the A- and D-genome loci was exploited in the design of genome-specific primers at each locus. The D-genome specificity of the primers was verified with Chinese Spring nullisomic-tetrasomic lines.

Primers amplifying products specific for the D genome were subsequently used to analyze sequence diversity in a total of 564 *T. aestivum* lines that included free-thresh-ing subspecies *aestivum, compactum,* and *sphaerococcum* and hulled subspecies *macha, petropavlovskyi, spelta, tibetanum, vavilovii,* and *yunnanense* together with 203 *Ae. tauschii* lines (see supplemental data at http://www.genetics.org/supplemental/).

Nucleotide diversity at the *Xwye838* locus: The amplified region of the D genome consisted of 114 bp of coding and 133 bp of noncoding sequence. For both *T. aestivum* and *Ae. tauschii*, no indels were detected and the only SNP observed within this region was a silent T-to-G transversion in the noncoding sequence. The *T. aestivum* accessions sampled are characterized by a vast preponderance of the *T* allele (frequency 0.984) with the *G* allele detected in 7 of 79 lines of bread wheat (ssp. *aestivum*) from Iran and in 1 of 13 lines of club wheat (ssp. *compactum*) from Afghanistan. In contrast, *Ae. tauschii* was characterized by a preponderance of the *G* allele (frequency 0.948) with similar frequencies in both the *Ae. tauschii* ssp. *tauschii* and ssp. *strangulata* genepools. The *T* allele was detected in two *Ae. tauschii* ssp. *strangulata* accessions and seven ssp. *tauschii* accessions. The presence of only two haplotypes in the *Ae. tauschii* germplasm was unexpected and demonstrated a surprisingly low level of diversity ($\theta = 0.00086$; Table 3).

Nucleotide diversity at the *Gss* locus: To assay SNPs at *Gss*, the D-genome sequence of a 326-bp region consisting of 117 bp of coding, 51 bp of noncoding, and 158 bp of the 3′ untranslated region was determined. The overall organization of sequence polymorphism in *Gss* is shown in Figure 1. The polymorphism can be grouped into 12 distinct haplotypes, 11 of them present in *Ae. tauschii* (Table 1). Seven indels were located in either the noncoding or 3′ untranslated regions with

Haplotype	*Ae. tauschii*	*T. aestivum*	*Ae. cylindrica*
1	0.217	0.985	0.043
2	0.010	0.000	0.000
3	0.364	0.000	0.000
4	0.005	0.000	0.000
5	0.035	0.000	0.000
6	0.000	0.015	0.000
7	0.131	0.000	0.000
8	0.035	0.000	0.000
9	0.005	0.000	0.000
10	0.187	0.000	0.957
11	0.005	0.000	0.000
12	0.005	0.000	0.000

an overall frequency of 1 indel/46 bp. The frequency of SNPs was 1 every 9 bp with 57% transitions and 35% transversions and the remaining 8% at positions where both transitions and transversions occurred. Of the 17 SNPs detected in the coding sequence, 7 resulted in five nonsynonymous substitutions (Figure 1). Two distinct haplotypes (*1* and *6*) were found among the 528 lines of *T. aestivum.* Haplotype *1* predominates and is found at a moderate frequency in *Ae. tauschii* and *Ae. cylindrica* (Table 1), whereas haplotype *6* was detected in only 7 of the 70 accessions of bread wheat sampled from Iran; this haplotype is very distinct from all the others, with 12 unique variants (Figure 2). Although we sampled 198 *Ae. tauschii* accessions, we did not detect haplotype *6.* On the basis of this sampling strategy the maximum population frequency of haplotype *6* that could occur in *Ae. tauschii* at a 95% confidence interval is 0.015.

To determine whether haplotype *6* was contributed by *Ae. cylindrica* via *T. aestivum* × *Ae. cylindrica* hybridization, *Gss* haplotype variation was assessed in 70 lines of *Ae. cylindrica* (see supplemental data at http://www.genetics.org/supplemental/). Only haplotypes *1* and *10* were found, the latter being most frequent. The maximum population frequency for haplotype *6* to escape detection ($P < 0.05$) in a sample size of 70 accessions is 0.042.

Estimates of nucleotide polymorphism, π and θ, are shown separately for the D genome of *T. aestivum* and *Ae. tauschii* (Table 2). These estimates are based on 282 bp of sequence data at the *Gss* locus and reveal that *Ae. tauschii* contains a 30-fold higher level of nucleotide polymorphism than does the hexaploid wheat D genome. Information about the selection history at the *Gss* locus can be obtained by applying Tajima's *D* statistic. The estimates of *D* were -1.846 for *T. aestivum* ($P < 0.05$) and 1.953 for *Ae. tauschii* ($0.1 > P > 0.05$). These estimates are indicative of an excess of rare variants in *T. aestivum* that were created by a second independent

hybridization event or introgression of D-genome material during polyploidization. The significant negative value for *D* in *T. aestivum* is, therefore, consistent with a recent origin (Dvorak *et al.* 1998a) of hexaploid wheat in which a variant arising by sexual hybridization was quickly fixed in the population.

Genetic relationships between taxa based on sequence polymorphism: A minimum distance-spanning tree (Digby and Kempton 1987) was constructed from the alignment of the 12 *Gss* haplotypes (Figure 2). These 12 haplotypes reveal four different groups of extant haplotypes in which the most recently derived haplotypes are also the most distinct and exist in the lowest frequency. This is illustrated by the cluster containing haplotypes *1–5* (Figure 2). Both proposed ancestral haplotypes, X and Y, are absent from our sample as well as several of the intermediate haplotypes that would correspond to each individual nucleotide substitution. Hexaploid wheat is represented by 2 of the 12 haplotypes (haplotypes *1* and *6*) and these are highly divergent.

DISCUSSION

The primary purpose of this study was to develop D-genome-specific PCR primers and to use them to determine the number of distinct lineages present in the allopolyploids *T. aestivum* and *Ae. cylindrica.* The organization of sequence polymorphism and haplotype content at two homeologous loci involved in wheat starch biosynthesis was determined and used to amplify D-genome-specific sequences.

Wheat D-genome origins: In the current study two sequence variants were encountered at each locus, suggesting that two genetically distinct progenitors have contributed to the D genome of *T. aestivum* accessions. For *Xwye838*, the SNP variant present in hexaploid wheat was also detected in *Ae. tauschii.* In contrast, only one of the two wheat *Gss* haplotypes was found in *Ae. tauschii*; wheat haplotype *6* was not detected in either *Ae. tauschii* or *Ae. cylindrica*, a weedy tetraploid goatgrass known to hybridize with wheat that contains an *Ae. tauschii* D genome (Gandilyan and Jaaska 1980; Snyder *et al.* 2000). Of the 11 *Gss* haplotypes found in *Ae. tauschii*, 2, haplotypes *1* and *10,* were encountered among 70 lines of *Ae. cylindrica* collected across the distribution of this species (Table 1). The frequency of haplotype *10* was 0.957 in *Ae. cylindrica* and 0.186 in *Ae. tauschii*; however, this haplotype was absent from the 564 lines of *T. aestivum* sampled.

Haplotype *6* theoretically could have been contributed by other Aegilops polyploid species with a D genome: *Ae. ventricosa, Ae. vavilovii, Ae. juvenalis,* and *Ae. crassa.* Although we did not investigate these species, none of them is a likely candidate on the basis of geographic distribution. Haplotype *6* was found in just a few of the wheat lines studied from Iran. Of the four Aegilops species, only *Ae. crassa* grows in Iran but it is not abun-
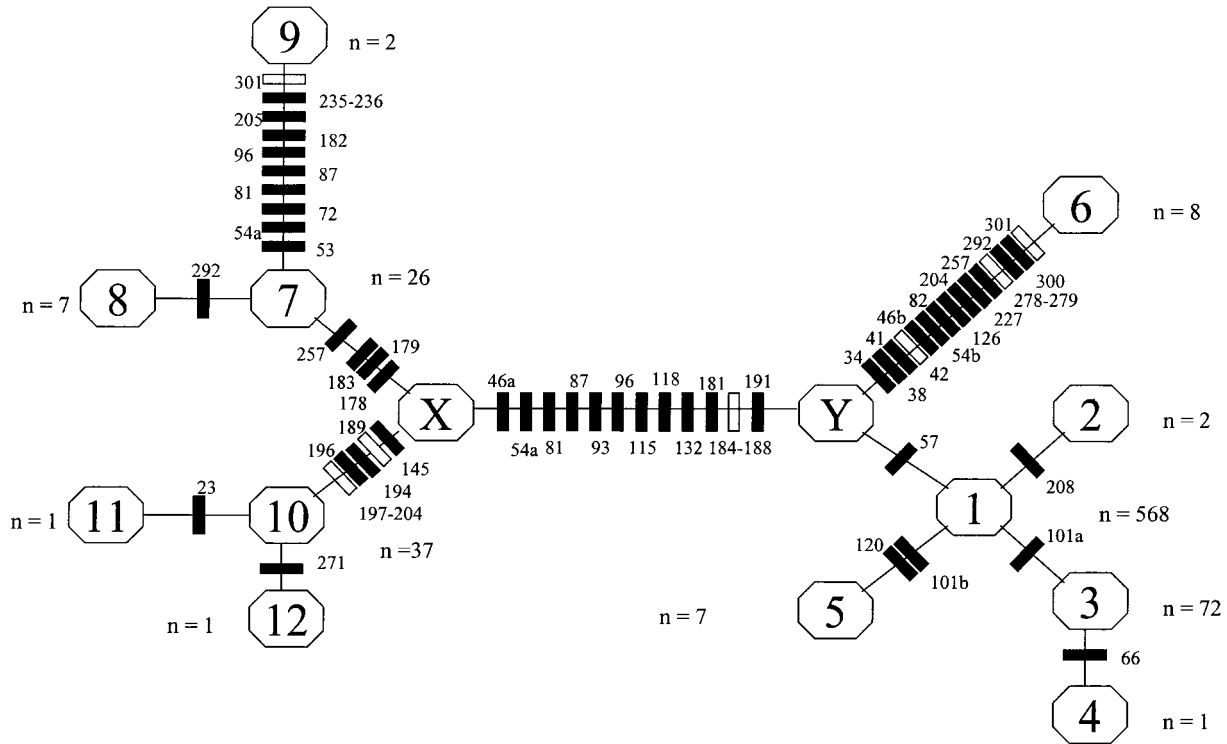
FIGURE 2.—A minimum-spanning tree was constructed from the compressed alignment. The polymorphisms that distinguish a given haplotype from another are located on the branches between them. Substitutions are indicated by rectangles: insertion/ deletions by an open rectangle and homoplastic substitutions by a solid rectangle.

dant there; the principal area of its distribution is Turkey, Kazakhstan, Uzbekistan, and Turkmenistan (VAN SLAGERN 1994). Haplotype *6* differs at 16 nucleotide positions (13 SNPs and three insertions/deletions) from haplotype *1* and differs from the corresponding A- and D-genome sequences detected in Chinese Spring by 20 substitutions and seven indels and 11 substitutions and four indels, respectively. Since *T. aestivum* is no more than 8000 years old (NESBITT and SAMUEL 1996), haplotype *6* cannot have evolved from haplotype *1* after the formation of wheat. Although we are unable to make

a comparison with the B genome, the likelihood of homeologous recombination generating haplotype *6* is low and it is more plausible that haplotype *6* was contributed to wheat by an *Ae. tauschii* population that was absent from our sample of 198 accessions.

At both loci, the polymorphism in wheat consisted of one common and one rare sequence. Both rare haplotypes were present in a limited number of wheat lines from Iran and a single accession from either Afghanistan or Turkey; only one accession harbored rare haplotypes for both *Xwye838* and *Gss*. The nature of the poly-

TABLE 2

Estimates of nucleotide diversity and test statistics for selection at *Gss* and *Xwye838*
in hexaploid wheat species and diploid *Ae. tauschii* species

	n	s	$\pi/(bp)$	$\theta/(bp)$	$\pi/(bp)_s$	$\theta/(bp)_s$	$\pi/(bp)_n$	$\theta/(bp)_n$	D
Gss									
Hexaploid (6x)	528	13	0.00135	0.00760	0.00206	0.001160	0.00000	0.00000	−1.84591*
Ae. tauschi (2x)	198	29	0.03558	0.02106	0.04676	0.02709	0.01501	0.00996	1.95293**
Xwye838									
Hexaploid (6x)	550	1	0.00023	0.00102	0.00037	0.00168	0.00000	0.00000	−0.68129
Ae. tauschii (2x)	173	1	0.00049	0.00086	0.00085	0.00150	0.00000	0.00000	−0.44740

All statistics were calculated without introns where n is the number of sequences, s is the number of segregating sites, π is the average pairwise diversity, θ is the diversity based on the number of segregating sites, and D is Tajima's test for neutrality. * and ** correspond to significance of $P < 0.05$ and $0.10 > P > 0.05$, respectively. Subscript s and n correspond to silent and nonsynonymous sites, respectively.

morphism detected at both loci in *T. aestivum* suggests that this polymorphism originated relatively recently via a tetraploid wheat × *Ae. tauschii* hybridization event rather than via mutation. Gene flow from *Ae. tauschii* into *T. aestivum* via hexaploid wheat × *Ae. tauschii* hybrids is unlikely (Dvorak *et al.* 1998a,b). On the basis of previous RFLP analysis (Dvorak *et al.* 1998a,b), the haplotype data obtained here, and the highly self-pollinating, homozygous nature of the species being studied, one can reasonably conclude that at least two *Ae. tauschii* sources contributed germplasm to the D genome of *T. aestivum*. The present study also establishes that the genepool of the D genome in *Ae. cylindrica* must also be a product of multiple hybridizations involving *Ae. tauschii* since both haplotypes detected in *Ae. cylindrica* are present in *Ae. tauschii*. This recurrent pattern of polyploidization has also been demonstrated for *Ae. triuncialis* ($2n = 4x$, genomes *UUCC*) (Wang *et al.* 1997) and suggests that this mode of origin may be a widespread phenomenon in the Aegilops-Triticum alliance.

Domestication bottlenecks: Domestication in most crop plants is associated with population bottlenecks and a reduction in genetic diversity. Small initial population sizes together with intense selection for agronomic traits are thought to have contributed to the narrow genetic base of most crop plants (Tanksley and McCouch 1997). Comparisons of DNA sequence variation between and within species are now becoming available and are providing new insights into the extent of divergence between species and their ancestors, particularly in maize (Hilton and Gaut 1998; White and Doebley 1999; Tenallion *et al.* 2001; Buckler and Thornsberry 2002). On the basis of two apparently neutral genes, *glb1* and *ahd1*, maize contains between 60 and 83% of the sequence diversity detected in progenitor species (Hilton and Gaut 1998). Similarly domesticated pearl millet contains ~67% of the sequence diversity found in its wild progenitors (Gaut and Clegg 1993). Tiffin and Gaut (2001) compared sequence diversity at four nuclear loci from autotetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*. No strong evidence for greater diversity in the tetraploid was obtained. On the basis of our *Gss* locus sequences, genetic diversity of the D genome of hexaploid wheat is, however, greatly reduced compared to that of *Ae. tauschii*. The nucleotide sequence diversity (π) was 30-fold higher in *Ae. tauschii* than in the *T. aestivum* D genome, and Tajima's *D* suggests a departure from the equilibrium neutral model at this locus, with an excess of rare sequence variants in hexaploid wheat species. This is compatible with a genetic bottleneck created by recent polyploidization. In contrast, the high diversity and positive value of *D* (Table 2) obtained for the diploid *Ae. tauschii* indicates an excess of intermediate frequency alleles, suggesting population subdivision in this species. This is compatible with the possibility (above)

that the *Gss* haplotype *6* may be present in some *Ae. tauschii* populations that were not sampled.

To our knowledge, this study represents the most extensive sequence-based haplotype analysis conducted to date for any member of the Triticeae. The only other published study of nucleotide diversity at homeologous loci is from allotetraploid Gossypium (Small *et al.* 1999), which identified low levels of diversity, reflecting a history of severe bottlenecks associated with speciation and recent domestication. The availability and exploitation of genome-specific amplicons together with sequence analysis will provide further impetus to investigate the evolutionary dynamics of speciation (Brumfield *et al.* 2003) and the mode of polyploid formation in plants such as wheat (Blake *et al.* 1999). For example, the methods outlined here for the D genome can be extended and used to investigate the structure of genetic variation in the wheat genome, particularly in relation to the number of genetically divergent lineages involved in the formation of AABB tetraploid wheat.

LITERATURE CITED

Ainsworth, C., F. Hosein, M. Tarvis, F. Weir, M. Burell *et al.*, 1995 Adenosine diphosphate glucose pyrophosphorylase genes in wheat: differential expression and gene mapping. Planta **197:** 1–10.

Blake, N. K., B. R. Lehfeldt, M. Lavin and L. E. Talbert, 1999 Phylogenetic reconstruction based on low copy DNA sequence data in an allopolyploid: the B genome of wheat. Genome **42:** 351–360.

Brumfield, R. T., P. Beeri, D. A. Nickerson and S. V. Edwards, 2003 The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. **18** (5): 249–256.

Buckler, E. S., and J. M. Thornsberry, 2002 Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. **5** (2): 107–111.

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

Digby, P. G. N., and R. A. Kempton, 1987 *Multivariate Analysis of Ecological Communities*. Chapman & Hall, London/New York.

Dvorak, J., P. E. McGuire and B. Cassidy, 1988 Apparent sources of the A genomes of wheat inferred from the polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome **30:** 680–689.

Dvorak, J., M.-C. Luo and Z.-L. Yang, 1998a Genetic evidence on the origin of *T. aestivum* L., pp. 235–251 in *The Origins of Agriculture and the Domestication of Crop Plants in the Near East*, edited by A. Damania. ICARDA, Aleppo, Syria.

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

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. USA **95:** 4441–4446.

Filatov, D. A., and D. Charlesworth, 1999 DNA polymorphism, haplotype structure and balancing selection in the *Leavenworthia PgiC* locus. Genetics **153:** 1423–1434.

Gandilyan, P. A., and V. E. Jaaska, 1980 A stable introgressive hybrid from hybridization between *Aegilops cylindrica* host and *Triticum aestivum* L. Genetika **16:** 1052–1058.

Gaut, B. S., and M. T. Clegg, 1993 Nucleotide polymorphism in the *Adh1* locus of pearl millet (*Pennisetum glaucum*) (*Poaceae*). Genetics **135:** 1091–1097.

Guadagnuolo, R., D. Savova-Bianchi and F. Felber, 2001 Gene

flow from wheat (*Triticum aestivum* L.) to jointed goatgrass (*Aegilops cylindrica* Host), as revealed by RAPD and microsatellite markers. Theor. Appl. Genet. **103:** 1–8.

HEY, J., 1998 Population genetics and human origins—haplotypes are key! Trends Genet. **14:** 303–304.

HILTON, H., and B. S. GAUT, 1998 Speciation and domestication in maize and its wild relatives: evidence from the *Globulin-1* gene. Genetics **150:** 863–872.

JAASKA, V., 1978 NADP-dependent aromatic alcohol dehydrogenase in polyploid wheat and their relatives. On the origin and phylogeny of polyploid wheat. Theor. Appl. Genet. **53:** 209–217.

JAASKA, V., 1981 Aspartate aminotransferase and alcohol dehydrogenase isozymes: intraspecific differentiation in *Aegilops tauschii* and the origin of the D genome polyploids in the wheat group. Plant Syst. Evol. **137:** 259–273.

KIHARA, H., 1931 Genomanalyse bit Triticum und Aegilops. II. Aegilotricum und Aegilops cylindrica. Cytologia **2:** 106–156.

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

LEITCH, I. J., and M. D. BENNETT, 1997 Polyploidy in angiosperms. Trends Plant Sci. **2:** 470–476.

LEVIN, D. A., 1980 Polyploidy and distribution, pp. 45–60 in *Polyploidy: Biological Relevance*, edited by W. H. LEWIS. Plenum, New York.

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

McFADDEN, E. S., and E. R. SEARS, 1946 The origin of *Triticum spelta* and its free-threshing hexaploid relatives. J. Hered. **37:** 81–89.

NEI, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

NESBITT, M., and D. SAMUEL, 1996 From staple crop to extinction? The archaeology and history of hulled wheat, pp. 41–100 in *Hulled Wheat: Promoting the Conservation and Use of Underutilized and Neglected Crops*, edited by S. PADULOSI, K. HAMMER and J. HELLER. International Plant Genetic Resources Institute, Rome.

NISHIKAWA, K., 1973 Alpha-amylase isozymes and phylogeny of hexaploid wheat, pp. 851–855 in *The 4th International Wheat Genetics Symposium*, edited by E. R. SEARS and L. M. S. SEARS. Missouri Agricultural Experiment Station, Columbia, MO.

NISHIKAWA, K., Y. FURUTA and T. WADA, 1980 Genetic studies on alpha-amylase isozymes in wheat. III. Intraspecific variation in *Aegilops squarrosa* and birthplace of hexaploid wheat. Japn. J. Genet. **55:** 325–336.

OSBORN, T. C., J. C. PIRES, J. A. BIRCHLER, D. L. AUGER, Z. J. CHEN *et al.*, 2003 Understanding mechanisms of novel gene expression in polyploids. Trends Genet. **19:** 141–147.

OTTO, S. P., and J. WHITTON, 2000 Polyploid incidence and evolution. Annu. Rev. Genet. **34:** 401–437.

PEBUSQUE, M.-J., F. COULIER, D. BIRNBAUM and P. PONTAROTTI, 1998 Ancient large-scale genome duplications: phylogenetic and linkage analyses shed light on chordate genome evolution. Mol. Biol. Evol. **15:** 1145–1159.

RAFALSKI, A., 2002 Application of single nucleotide polymorphisms in crop genetics. Curr. Opin. Plant Biol. **5:** 94–100.

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

SALAMINI, F., H. OZKAN, A. BRANDOLINI, R. SCHÄFER-PREGL and W. MARTIN, 2002 Genetics and geography of wild cereal domestication in the Near East. Nat. Rev. Genet. **3:** 429–441.

SEEFELDT, S. S., R. ZEMETRA, F. L. YOUNG and S. S. JONES, 1998 Production of herbicide-resistant jointed goatgrass (*Aegilops cylindrica*) × wheat (*Triticum aestivum*) hybrids in the field by natural hybridization. Weed Sci. **46:** 632–634.

SHARIFLOU, M. R., and P. J. SHARP, 1999 A polymorphic microsatellite in the 3′ end of 'waxy' genes of wheat, *Triticum aestivum.* Plant Breed. **118:** 275–277.

SMALL, R. L., J. A. RYBURN and J. F. WENDEL, 1999 Low levels of nucleotide diversity at homeologous *Adh* loci in allotetraploid cotton (*Gossypium* L.). Mol. Biol. Evol. **16:** 491–501.

SNYDER, J. R., C. A. MALLORY-SMITH, S. BALTER, J. L. HANSEN and R. S. ZEMETRA, 2000 Seed production on *Triticum aestivum* by *Aegilops cylindrica* hybrids in the field. Weed Sci. **48:** 588–593.

SOLTIS, D. E., and P. S. SOLTIS, 1993 Molecular data and the dynamic nature of polyploidy. Crit. Rev. Plant Sci. **12:** 243–273.

SOLTIS, D. E., and P. S. SOLTIS, 1999 Polyploidy: recurrent formation and genome evolution. Trends Ecol. Evol. **14:** 348–352.

STEBBINS, G. L., 1947 Types of polyploids: their classification and significance. Adv. Genet. **1:** 403–429.

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

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.

TALBERT, L. E., L. Y. SMITH and N. K. BLAKE, 1998 More than one origin of hexaploid wheat is indicated by sequence comparison of low-copy DNA. Genome **41:** 402–407.

TANKSLEY, S., and S. McCOUCH, 1997 Seed banks and molecular maps: unlocking genetic potential from the wild. Science **277:** 1063–1066.

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

TIFFIN, P., and B. S. GAUT, 2001 Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. Genetics **158:** 401–412.

VAN SLAGERN, M. W., 1994 *Wild Wheat: A Monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae).* Wageningen Agricultural University Papers 94-7, Wageningen, The Netherlands.

WANG, G.-Z., N. MIYASHITA and K. TSUNEWAKI, 1997 Plasmon analyses of *Triticum* (wheat) and *Aegilops*: PCR-single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNA. Proc. Natl. Acad. Sci. USA **94:** 14570–14577.

WATTERSON, G. A., 1975 On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. **7:** 188–193.

WENDEL, J. F., 2000 Genome evolution in polyploids. Plant Mol. Biol. **42:** 225–249.

WHITE, S. E., and J. F. DOEBLEY, 1999 The molecular evolution of *terminal earl*, a regulatory gene in the genus Zea. Genetics **153:** 1455–1462.

WOLFE, K. H., and D. C. SHIELDS, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:** 708–713.