Chromosome Mapping and Phylogenetic Analysis of the Cytosolic Acetyl-CoA Carboxylase Loci in Wheat

Justin Faris,†[1] *Anchalee Sirikhachornkit,** *Robert Haselkorn,** *Bikram Gill,*† *and Piotr Gornicki**

*Department of Molecular Genetics and Cell Biology, University of Chicago; and †Department of Plant Pathology, Kansas State University

The cytosolic isoform of plant acetyl-CoA carboxylase is a multidomain enzyme involved in the synthesis of very-long-chain fatty acids and in secondary metabolism. Chromosome mapping of wheat identified one locus containing cytosolic acetyl-CoA carboxylase genes (*Acc-2*) and a related partially processed pseudogene (*Ψ-Acc-2*) in the distal region of the long arm of wheat homoeologous group 3 chromosomes. Multiple copies of the *Acc-2* genes, whose presence was suggested by sequence analysis, are likely to be arranged in tandem repeats. At least three out of five genes cloned from hexaploid wheat map to this locus. Another locus containing *Acc-2*–related sequences is present in the distal region of the long arm of chromosome 5D. The identity of the hybridizing DNA present at this locus remains unknown. A system based on PCR-cloning and DNA sequence analysis of acetyl-CoA carboxylase genes was developed to address various phylogenetic and systematics questions in grasses. It was applied to reconstruct the phylogeny of the *Acc-2* genes from D- and S-genome *Aegilops* and A-genome *Triticum* diploid species, AABB- and AAGG-genome tetraploid wheat, and AABBDD-genome hexaploid wheat, as well as from rye and barley. The combined cytogenetic and molecular evolution approach allowed assignment of gene sequences included in phylogenetic analysis to specific loci on homoeologous chromosomes. Recurring gene duplication followed by chromosome translocation and/or possible loss of some gene copies, as well as loss of introns, occurred in the gene family in different plant lineages. Two major *Acc-2* clades appeared before the divergence of barley and rye. Nucleotide substitution rates in different parts of the *Acc-2* gene were assessed. This analysis of the *Acc-2* loci provides detailed information regarding evolutionary events at a low–copy-number locus containing important functional genes. These events are likely to be common and to play a significant role in shaping grass genomes.

Introduction

To understand how genes evolve in polyploid species, two factors need to be considered: the origin and history of individual genomes, and the structure of the gene family being investigated. The results of such analysis may provide important information on genome-shaping events and processes occurring at specific loci containing functional genes and contribute to our understanding of functional aspects of these genes, as well as phylogenetic relationships between species. Cytogenetic, molecular evolution, and molecular biology approaches were combined to address these problems. The study presented in this paper focused on a small gene family encoding cytosolic acetyl-CoA carboxylase (ACCase) in wheat and related grasses.

All members of the Triticeae tribe, including wheat, barley, and rye, are presumed to have evolved from a common ancestor. Among them are polyploid wheat species such as the tetraploid durum (macaroni) wheat *Triticum turgidum* and the hexaploid bread wheat *Triticum aestivum*, which arose through processes of interspecific hybridization events among diverged ancestral wheats followed by spontaneous chromosome doubling. These polyploids contain entire genomes of two or three species in the homozygous condition. The major genomes of wheat and its relatives, named A, B, D, G, S, etc., have a basic chromosome number of seven and are homoeologous (reflecting residual homology of originally completely homologous chromosomes) to one another. The Triticeae tribe presents an interesting case because it includes species of different ploidy for which some key evolutionary steps are known.

Genetic studies have revealed that the polyploid wheat species constitute two evolutionary lineages (reviewed in Cox 1998). *Triticum turgidum* (AABB genome) and *T. aestivum* (AABBDD) compose one lineage, and *Triticum timopheevii* (AAGG) and *Triticum zhukovskyi* (AAAAGG) compose the other lineage. The wild tetraploids *T. turgidum* ssp. *dicoccoides* (AABB) and *T. timopheevii* ssp. *armeniacum* (AAGG) arose as amphiploids from the hybridization between *Triticum urartu* (AA) (Dvorak et al. 1993; Takumi et al. 1993) and two different plasmon (the sum of extrachromosomal hereditary determinants) types of another wild diploid wheat which is thought to be *Aegilops speltoides* (SS) (Dvorak and Zhang 1990; Wang, Miyashita, and Tsunewaki 1997). *Triticum aestivum* (AABBDD) arose from spontaneous hybridization between *T. turgidum* (AABB) and the diploid goatgrass *Aegilops tauschii* (DD) (Kihara 1944; McFadden and Sears 1946; reviewed in Dvorak et al. 1998). In *T. zhukovskyi* (AAAAGG), one set of A genomes was contributed by *T. urartu* and the other was contributed by *Triticum*

[1] Present address: USDA/ARS Cereal Crops Research Unit, Northern Crop Science Laboratory, Fargo, North Dakota.

Abbreviations: ACCase, acetyl-CoA carboxylase; *Acc-2*, gene encoding cytosolic ACCase; *Ψ-Acc-2*, *Acc-2* related partially processed pseudogene; dt line, ditelosomic line; NT line, nullitetrasomic line; RIL, recombinant inbred line.

Key words: Triticeae, *Aegilops*, pseudogene, gene sequence, grass, evolution, plant.

Address for correspondence and research: Piotr Gornicki, Department of Molecular Genetics and Cell Biology, University of Chicago, 920 East 58th Street, Chicago, Illinois 60637. E-mail: pg13@midway.uchicago.edu.

Mol. Biol. Evol. 18(9):1720–1733. 2001

Table 1
PCR Primers

		Forward Primer	Reverse Primer
Acc-2,1 verification	First step	TTCATCTCCCACACATAACACGAA	TGGTCAGATTCCACCATTATTGCC
	Second step	CATTCTAGTCCAACATTCATGGAG	TTAGTCCAGGATGATAGGATTCTG
Acc-2,2 verification	First step	CAAAAGCATGATATGCCCTTGTGGC	TGGTCAGATTCCACCATTATTGCC
	Second step	TTTTTGAGGATGCAATGGTGCAC	AGAGAAGATATGTTTTCAGCCGAG
Acc-2,3 cloning		TCCGGCCGAACCGACGGTACGCGC	TGGTCAGATTCCACCATTATTGCC
Acc-2,4 cloning		CAAAAGCATGATATGCCCTTGTGGC	TGGTCAGATTCCACCATTATTGCC
Acc-2,5 cloning		TGGCGGCGCGCCTCCGGACGGACC	TGGTCAGATTCCACCATTATTGCC
Probe 5′F11 (0.75 kb)		CCAGCGGGCCATGTCACTACC	AACAGCTATGACCATG (vector primer)
Probe Intl (0.36 kb)		TGGCGGCGCGCCTCCGGACGGACC	CAGACGGGGCGAACCCGGCAATCC
Probe Int2 (0.44 kb)		TCCGGCCGAACCGACGGTACGCGC	CAGACGGCACGAAGAAGCCGCCCG
Probe Pro1 (0.30 kb)		GATTAAATCATTCGCTCCAGAACT	CATTCTAGTCCAACATTCATGGAG
Probe Pro2 (0.36 kb)		CAAGGGGAAATGGAATCGACTCCG	TCAAGTGTATGGACATACACGCGC
Probe Pro21 (0.40 kb)		TTCTTATTTGATTGTTTAATAGTA	TTTTTGAGGATGCAATGGTGCAC
Probe Pro4 (0.29 kb)		CCCTTGTGGCGCAACCAGTGACAC	GTAGGTAGGGCCCCAACGCCTTGG
Acc-2 cloning for phylogenetic analysis		GTCCCCGGATCGCCTATATTTATT	TTCAAGAGATCCACRGTGTAGTCA
		CTATATTTATTATGAAGGTGGCATC	AGATCCACRGTGTAGTCAACATTA
			ATTTGGATCTCTTTTAACCCAAGTA
Ψ-Acc-2,2 cloning		CTACAGCTTATGGACGGATTGGTCTA	CATGTATCCGGACACTACCCCTTAC

monococcum, which is a close relative of *T. urartu* (Dvorak et al. 1993). Therefore, *T. zhukovskyi* originated from the hybridization of *T. timopheevii* (AAGG) with *T. monococcum* (AA) to complete the second lineage (Upadhya and Swaminathan 1963).

The homoeologous wheat chromosomes are, for the most part, colinear and composed of the same genes; e.g., homoeologous chromosomes 1A, 1B, and 1D are presumed to contain the same genes at equivalent positions along the chromosomes. This chromosome colinearity extends to some degree to the entire grass family (Bennetzen et al. 1998; Gale and Devos 1998). Polyploidization and large-scale chromosome rearrangements, evolutionary events, and processes occurring at low–copy-number loci containing functional genes, as well as changes in intergenic regions work in parallel to shape the grass genomes. Gene duplications followed by evolution of new tissue and development specificity or subcellular targeting add plasticity to the system. Transposable elements play an important role in these processes. The contribution of these various phenomena to evolution of plant genomes and genes has recently been discussed (Gaut 1998; Bennetzen 2000; Fedoroff 2000; Wendel 2000).

A significant amount of information about ACCase and its genes and their structure and function is already available from other studies (Gornicki and Haselkorn 1993; Gornicki et al. 1994, 1997; Podkowinski et al. 1996; unpublished results). The cytosolic isoform of ACCase present in plants provides malonyl-CoA for the synthesis of very–long-chain fatty acids and for secondary metabolism. It is a multidomain enzyme of eukaryotic origin, as indicated by its subunit structure and by amino acid sequence comparisons (Gornicki et al. 1994, 1997; Konishi et al. 1996; Podkowinski et al. 1996; Incledon and Hall 1997). A separate ACCase isozyme provides malonyl-CoA for de novo fatty acid biosynthesis in the plastid. The latter is a multisubunit enzyme of prokaryotic (endosymbiont) origin in most plants other than Poaceae and a multidomain enzyme in Poaceae (Egli et al. 1995; Konishi et al. 1996; Gornicki et al. 1997; Schulte et al. 1997; Christopher and Holtum 2000). The genes encoding wheat cytosolic and plastid ACCase are ~15 kb in size and have a large number of introns. The two proteins are 2,260 and 2,311 amino acids long, respectively, and their sequences are 67% identical. A putative plastid transit peptide is present at the N-terminus of the plastid isozyme (Gornicki et al. 1997).

In this paper, we describe the chromosome localization of the cytosolic ACCase genes (*Acc-2*) and a related pseudogene (*Ψ-Acc-2*) in hexaploid wheat (*T. aestivum*), as well as phylogenetic relationships for the genes in wheat and some other representative species of the Triticeae tribe. The nature of evolutionary changes in different parts of the *Acc-2* gene are also discussed.

Materials and Methods

Plant Materials

Seeds used in this study were obtained from the Wheat Genetics Resource Center (Kansas State University, Manhattan, Kans.).

Cloning and Analysis of *Acc-2* Genes

Identity of the 5′ ends of genomic clones 191 (*Acc-2,1*, GenBank AF305204 and U39321) and 153 (*Acc-2,2*, GenBank AF305205), isolated previously (Podkowinski et al. 1996), was confirmed in two steps by PCR using primers shown in table 1. Fragments of three new genes, *Acc-2,3* (GenBank AF305206), *Acc-2,4* (GenBank AF305207), and *Acc-2,5* (GenBank AF305208) were cloned by PCR using primers shown in table 1. The fragment of *Acc-2,3* included the first intron and

short fragments of the flanking exons. The fragment of *Acc-2,4* included the 5′-end fragment of the gene, the first intron, and a short fragment of exon 2. The fragment of *Acc-2,5* included the first intron and a short fragment of exon 2. A 5-kb fragment of genomic clone 232 containing a putative pseudogene (*Ψ-Acc-2,1*) isolated previously (Podkowinski et al. 1996) was sequenced (GenBank AF305209). A 0.4-kb fragment of another copy of the pseudogene (*Ψ-Acc-2,2*) was cloned by PCR using primers shown in table 1 and sequenced (GenBank AF362956). Wheat (*T. aestivum* cv. Tam 107) genomic DNA (Clontech) was used as PCR template. PCR products were cloned into the vector pC2.1 (Invitrogen) and sequenced.

Cytogenetic Analysis of *Acc-2* Genes in Wheat

The polyploid nature of *T. aestivum* allows it to tolerate, and transmit through gametes, a certain degree of aneuploidy, i.e., an abnormal chromosome constitution. Wheat aneuploid stocks including nullisomic-tetrasomic (NT) (Sears 1944; Sears 1966), ditelosomic (dt) (Sears and Sears 1978), and chromosome deletion lines (Endo and Gill 1996) have been developed for use as genetic tools. These stocks allow assignment of genes to genomes, chromosomes, and chromosome arms. NT lines are missing a pair of chromosomes (nullisomic) but have an extra pair of a homoeologous chromosome (tetrasomic); e.g., a stock that is N3B T3D would be missing the pair of 3B chromosomes but would have two pairs of 3D chromosomes to compensate for the missing pair. NT stocks are therefore used for assigning genes and molecular markers to individual chromosomes. Dt lines lack a pair of chromosome arms. For example, the line dt3AS has only the short arms of the pair of 3A chromosomes present, and it lacks the 3A long arms. Dt lines are therefore very useful for assigning genes and molecular markers to chromosome arms.

NT and dt lines of Chinese Spring wheat were used in this study to assign *Acc-2* genes to individual chromosomes and chromosome arms, respectively. N2AT2B and N4BT4D plants were identified cytologically, as these stocks were maintained as monosomic-tetrasomic lines. Dt lines 2AL, 4AS, 5AS, 2BS, 4BL, 5BS, and 5DS were cytologically identified in the progeny of ditelo-monotelo lines.

To take the cytogenetic analysis one step further, genes and markers can be assigned to chromosome arm regions using the chromosome deletion lines. These lines have terminal chromosome deletions with breakpoints occurring at various positions along the chromosome. An approximate physical location of a molecular marker on a particular chromosome arm is indicated by the presence or absence of a restriction fragment detected by hybridization of a specific DNA probe to DNA from lines that each have different but known sizes of terminal chromosome deletions. Over 400 lines having terminal chromosome deletions of various sizes have been isolated (Endo and Gill 1996) and used to construct physical maps of the wheat chromosomes (Gill, Gill, and Endo 1993; Hohmann et al. 1994; Delaney et al. 1995*a*, 1995*b*; Mickelson-Young, Endo, and Gill 1995; Gill et al. 1996*a*, 1996*b*). For this study, chromosome deletion stocks having terminal deletions in the long arms of groups 3 and 5 were used.

The physical mapping experiments were complemented by placing *Acc-2* genes on existing genetic linkage maps (GrainGenes, wheat.pw.usda.gov) using 135 recombinant inbred lines (RILs) derived from a cross between a synthetic hexaploid wheat, W-7984, and the common hexaploid wheat variety Opata 85 (Nelson et al. 1995). The population was provided by Dr. M. E. Sorrells (Cornell University, Ithaca, N.Y.). *Acc-2* genes were also placed on the genetic map of the D genome using a mapping population consisting of 56 F_2 progeny derived from the cross of *A. tauschii* Coss. accessions TA1691, var. *meyeri*, and TA1704, var. *typica* (Kam-Morgan, Gill, and Muthukrishnan 1989; Gill et al. 1991). As described in the *Introduction*, *A. tauschii* is the D-genome progenitor of hexaploid bread wheat (*T. aestivum*), and its chromosomes are colinear with the corresponding D-genome chromosomes of hexaploid wheat.

Hybridization Probes

A series of cDNA hybridization probes was used for Southern analysis. First, cDNA probes ucg21 and ucg22, corresponding to the coding part of *Acc-2* genes, were used to simultaneously detect multiple gene copies and to identify all chromosome loci which contained them. These probes were expected to hybridize with all cytosolic ACCase genes whose coding sequences were ~98% identical (Gornicki et al. 1994; Podkowinski et al. 1996), as well as with the putative pseudogene (*Ψ-Acc-2*), whose sequences (GenBank AF305209 and AF362956) were more than 90% identical to the *Acc-2* coding sequence. Second, gene-specific probes derived from less-conserved regions of the *Acc-2* genes, the 5′-end portion and the first intron, were used to identify and map individual genes. Probe ucg21 is a 7-kb *Sal*I fragment, and ucg22 is a 3.4-kb *Bam*HI fragment of the full-length wheat cytosolic ACCase cDNA described previously (Joachimiak et al. 1997). 5′Fl2 is a 0.53-kb *Sal*I-*Sac*II fragment of the *Acc-2,2* lambda genomic clone 153. Other probes were prepared by PCR using cloned fragments of genomic DNA as templates. PCR primers and probe sizes are shown in table 1. The approximate positions of the target sites for these hybridization probes are shown in figure 1.

Southern Analysis

Isolation of DNA, gel electrophoresis, restriction enzyme digestion, probe labeling, Southern blotting, DNA hybridization, and membrane washing procedures were performed as described previously (Faris, Haen, and Gill 2000). Probes representing *Acc-2* genes were hybridized to Southern blots of NT and dt stocks, and resulting autoradiographs were visually scored to identify fragments absent in any of the stocks. If a fragment was absent in a particular NT stock, we inferred its location to be on the chromosome in the nullisomic con-

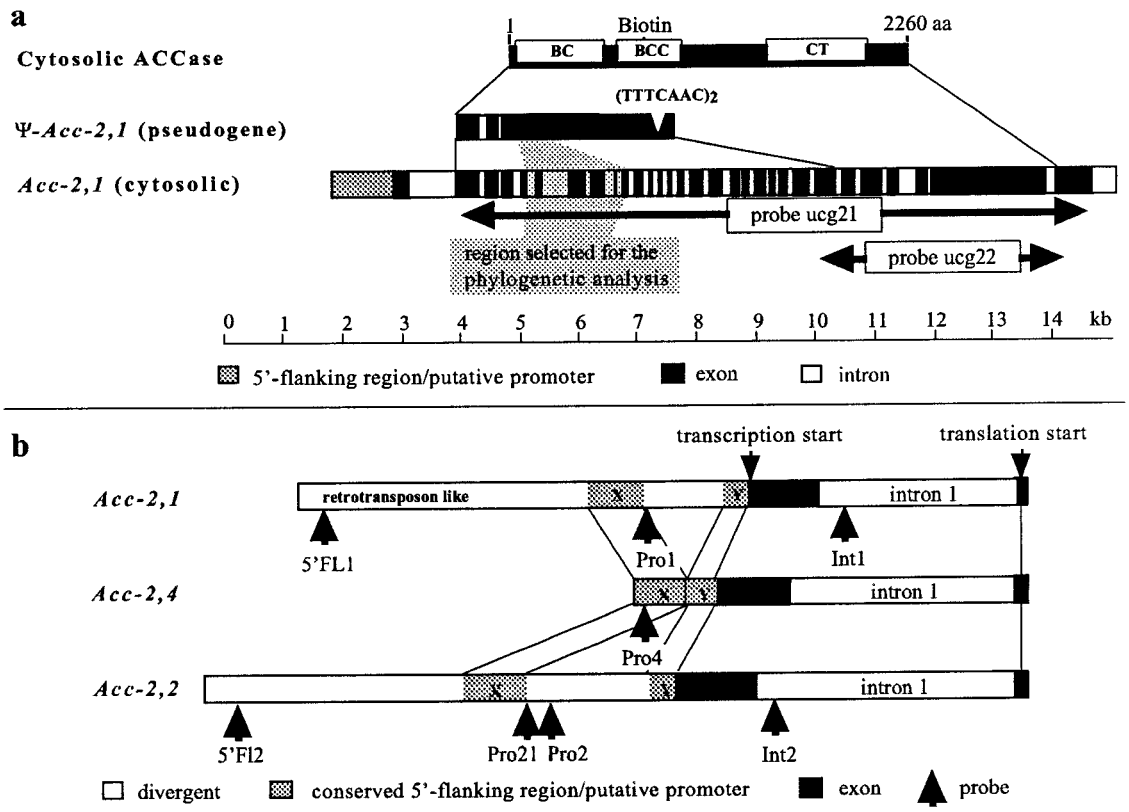

FIG. 1.—The structure of wheat cytosolic ACCase and its gene (*a*), and the structure of the 5′-end portion of three *Acc-2* genes (*b*). *a*, ACCase functional domains: biotin carboxylase (BC), biotin carboxyl carrier containing the biotin attachment site (BCC) and carboxyltransferase (CT), and intron-exon structure of the genes were described previously (Gornicki et al. 1994, 1997; Podkowinski et al. 1996). The gene fragment selected for the phylogenetic analysis is shaded in gray. cDNA probes ucg21 and ucg22 (genetic marker *Xucg2*, fig. 2) correspond to gene regions indicated by horizontal arrows and are defined in *Materials and Methods.* The approximate position of a 7-nt repeat found in the *Ψ-Acc-2,1* copy of the pseudogene is indicated by a white triangle. Sequences of *Acc-2,1* and *Ψ-Acc-2,1* were deposited in GenBank under accession numbers AF305204 and U39321, and AF305209, respectively. *b*, The 5′-end portion of the *Acc-2* genes consists of two blocks of conserved sequences (X and Y) separated by a highly divergent expansion element. Approximate locations of target sites for hybridization probes 5′FL1, 5′FL2, Pro1, Pro2, Pro21, Pro4, Int1, and Int2 are shown by arrowheads and are defined in *Materials and Methods.* The presence of a retrotransposon-like element at the 5′ end of *Acc-2,1* was suggested by sequence similarity. Sequences of the 5′-end portions of *Acc-2* genes were deposited in GenBank under accession numbers AF305204–AF305207.

dition. The concomitant presence of a double-dose fragment in the stocks tetrasomic for a particular chromosome was taken as additional evidence for the chromosome localization of fragments. In the analysis of dt lines, a fragment absent in a stock indicated its presence on the opposing arm of the chromosome. *Acc-2* gene-specific probes were used to survey for polymorphism between the parents of the mapping populations by hybridizing them to the parental DNA digested with five different restriction enzymes (*Eco*RI, *Eco*RV, *Dra*I, *Hin*dIII, and *Xba*I). These probes were then hybridized to DNA of the individuals of the populations digested with the enzyme giving the clearest polymorphism. Linkage analysis was done using MAPMAKER V2.0 (Lander et al. 1987) with a LOD of 2.0 and the Kosambi mapping function (Kosambi 1944).

Acc-2 Gene Sequences for Phylogenetic Analysis

Fragments of the *Acc-2* gene were cloned by High Fidelity PCR carried out according to the manufacturer's protocol (Roche) using different combinations of primers shown in table 1. All components of the PCR except the DNA polymerase were incubated for 2–3 min at 94°C. The PCR was then initiated by the addition of polymerase. Amplification was for 35 cycles of 0.5 min at 94°C, 0.5 min at 52°C, and 4–6 min at 68°C using 0.5–1.0 μg of genomic DNA as template in a 50-μl reaction. DNA from plant material was extracted as described previously (Faris, Haen, and Gill 2000). The PCR products were cloned into the vector pGEM-T Easy (Promega) and sequenced (University of Chicago Cancer Center Sequencing Facility). Vector primers, as well as a series of universal gene-specific primers, were used for sequencing.

PCR-cloning-based sequence analysis of genes from species of different ploidy, some of which were genetically very closely related, required strict quality control to eliminate cloning and sequencing artifacts. The analysis became even more complex for multicopy gene families. First, for each gene, multiple clones were sequenced and compared to eliminate PCR related errors. High-quality sequence of both strands was needed to resolve sequencing-related problems. Second, consensus sequences generated for all of the genes were ana-

lyzed in pairwise and multiple-sequence comparisons to verify single-nucleotide differences and to exclude sequences of chimeric DNA molecules. Third, a large number of clones representing the gene family from each species, obtained in independent PCR experiments using different pairs of primers, were analyzed to overcome any bias of PCR amplification and to assess the maximum number of gene copies. Finally, multiple-sequence alignments were created and then inspected by eye at all polymorphic sites to verify substitutions found in a single taxon. PCR primer target sites were excluded from the alignments.

Only two single-clone sequences were included in the phylogenetic analysis: an *Acc-2,2* sequence obtained from a clone isolated from a genomic library, and the barley (cyt2) sequence. The number of nucleotide differences between two barley sequences (cyt1 and cyt2) was much higher than could be accounted for by assuming only PCR errors, and there was no indication that the cyt2 clone was a chimera. The barley gene represented by cyt2 had two fewer introns than all other *Acc-2* genes included in the analysis.

Sequence and Phylogenetic Analysis Software

Sequencher (Gene Codes Corporation, Ann Arbor, Mich.) was used to manage the sequencing project. ClustalX (Thompson, Higgins, and Gibson 1994) was used to create multiple-sequence alignments. MacClade (W. P. Maddison and D. R. Maddison, University of Arizona, Tucson; phylogeny.arizona.edu/macclade/macclade) was used for analysis of multiple-sequence alignments. PAUP 4.0b5 (D. Swofford, Smithsonian Institution, Washington, D.C.; lms.si.edu/PAUP/) and MEGA (S. Kumar, K. Tamura, and M. Nei, Pennsylvania State University, University Park, Penn.; evolgen.biol.metro-u.ac.jp/MEGA) were used to calculate genetic distances and create phylogenetic trees.

Multiple-Sequence Alignment and Phylogenetic Analysis

After an initial phylogenetic analysis performed on all sequences at once to identify major clades, the final sequence alignment was created in four steps. First, all *Triticum*/*Aegilops* sequences of clade *Acc-2*I were aligned. Rye, barley, and wheat *Acc-2,2* sequences were then added to the alignment. A short $(CA)_n$ repetitive sequence found in *Acc-2,2* was treated as a single insertion rather than aligned with other C-rich sequences in this region, as suggested by the outcome of the ClustalX analysis. The six sequences of the clade *Acc-2*II were aligned separately. The two alignments were aligned with each other. At this stage, the alignment was adjusted by hand within variable segments of two introns. Finally, a *Lolium rigidum* sequence (cyt1, GenBank AF343454) was added to the alignment as an outgroup. Multiple alignment of all *Acc-2* sequences at once did not provide any clear suggestions of possible further improvements. No attempts were made to improve the alignment of the *Lolium* sequence, which remained problematic at some intron sites. The alignment was 2,081 nt long, with 696 exon positions and 1,439 invariant positions. The gene sequences were deposited in GenBank under accession numbers AF306803–AF306829. This alignment was separated into exon and intron parts. The intron-exon structure of the *Acc-2* gene had previously been established (Podkowinski et al. 1996). The alignment of exon sequences was unambiguous and without gaps. Multiple indels were found in introns. These alignments were used to create phylogenetic trees.

Phylogenetic trees were created in several different ways. First, phylogenetic trees based on intron sequences only were created by the Neighbor-Joining method excluding gaps from pairwise comparisons and without correction for multiple substitutions. This tree was characterized by very good support for major clades, indicated by bootstrap values >80% of 1,000 replicates. Second, Neighbor-Joining trees excluding gaps from pairwise comparisons but with correction for multiple substitutions by Jukes-Cantor methods and similar trees based on gene sequences (exons and introns) were generated. These trees had the same topology at most of those well-supported nodes. Finally, the same conclusion was reached for consensus trees generated by the heuristic maximum-parsimony search (equally weighted characters and nucleotide transformations, 1,000 random-addition replicates, tree bisection-reconnection branch swapping, gaps treated as missing data). Forty-two best trees (length 815) were found for the *Acc-2* gene (exons and introns) based on 310 informative characters (consistency index [CI] = 0.896; retention index [RI] = 0.935). Sixty-three best trees (length 668) were found for the *Acc-2* gene based on 242 informative characters in introns (CI = 0.910, RI = 0.941). Parsimony bootstrap analysis followed the same scheme, with 1,000 replicates, each with 10 random-addition replicates.

The alignment of the 5′-end portions of three *Acc-2* genes (*Acc-2,1*, *Acc-2,2*, and *Acc-2,4*; GenBank accession numbers AF305204, AF305205, and AF305207) was 2,361 bp long with 1,879 invariant positions and included sequences from the conserved block X to the fifth codon of the ACCase open reading frame located in the second exon of the gene (fig. 1*b*). The intron-exon structure of the 5′-end portion of the *Acc-2* gene had previously been established (Podkowinski et al. 1996).

Merged alignments of the coding region exon and intron sequences were used to calculate nucleotide substitution rates at synonymous positions and at all intron positions, respectively. In addition, substitution rates at all positions were calculated for individual introns, as well as some conserved segments of the 5′-end portion of the gene, based on their sequence alignments, described above. These rates were not corrected for multiple substitutions, and gaps were excluded only from pairwise comparisons.

Results

Chromosome Localization of the *Acc-2* Loci

The chromosome localization of the *Acc-2* genes was determined by hybridizing probe ucg21 and/or

ucg22 to genomic DNA from NT lines, dt lines, and chromosome deletion lines. NT analysis of probes ucg21 and ucg22 indicated that the *Acc-2* genes are located on homoeologous group 3 chromosomes (fig. 2*a*). These probes also detected a fragment on chromosome 5D (fig. 2*a*). Analysis of dt lines indicated that these probes hybridize to the long arms of group 3 chromosomes and chromosome 5D (fig. 2*b*). Hybridization of ucg21 to DNA of group 3 and group 5 chromosome deletion lines mapped the *Acc-2* genes near the distal ends of the long arms of group 3 chromosomes and chromosome 5D (fig. 2*c*). Positions of the *Acc-2* loci on wheat physical maps are shown in figure 3.

In the hexaploid wheat population of RILs, probe ucg21 detected a locus that mapped near the distal end of the long arm of chromosome 3A (fig. 3*a*), but fragments for 3B, 3D, and 5D were monomorphic and therefore could not be mapped in this population.

To distinguish whether the *Acc-2* locus on chromosome 5D arose before or after polyploidization, the *Acc-2* gene was mapped in *A. tauschii*, the diploid D genome progenitor of hexaploid wheat. The genetic position of the *Acc-2* genes was determined by hybridizing probe ucg21 to DNA of the segregating mapping populations (data not shown) and placing the detected loci on existing maps of *A. tauschii*. Linkage analysis of probe ucg21 in the diploid *A. tauschii* F_2 population located the site distally on 3DL and 5DL (fig. 3). These results suggested that the 5D *Acc-2* locus already existed in the diploid species.

Southern Analysis and Chromosome Mapping Using Gene-Specific Probes

NT and dt analysis using *Acc-2,1*-specific probes Pro1 and Int1 indicated that they both hybridize to common fragments on the long arms of homoeologous group 3 chromosomes (fig. 4*a*), but they did not detect any hybridizing DNA on chromosome 5D that would correspond to the locus detected by ucg21 and ucg22. Genetic linkage mapping of Pro1 and Int1 indicated that they cosegregate with each other and with ucg21 on the long arm of chromosome 3A in the RIL population and on 3DL in the *A. tauschii* F_2 population (fig. 3*b*). It is most likely that Pro1 and Int1 detect homoeoloci on chromosomes 3A, 3B, and 3D in hexaploid wheat, but the 3B and 3D fragments were monomorphic in the RIL population and therefore could not be located on the genetic map. Based on phylogenetic analysis of the coding region described later in this paper, the *Acc-2,1* gene was assigned to the A genome. All of this information indicates that a copy of the *Acc-2,1* gene is present on each of the group 3 chromosomes of genomes A, B, and D.

NT and dt analysis using *Acc-2,2*–specific probe Pro21 and *Acc-2,4*–specific probe Pro4 (fig. 4*b*) indicated that these probes hybridize to the same fragments on the long arms of chromosomes 3B and 3D, but not chromosome 3A or 5D. Genetic linkage analysis of probes Pro21 and Pro4 revealed that they cosegregate with each other and with ucg21, Pro1, and Int1 on the long arm of chromosome 3D in *A. tauschii* (fig. 3*b*). Pro21 and Pro4 did not detect polymorphisms between W7984 and Opata 85 and therefore were not mapped in the RIL population. These data indicate that a copy of the *Acc-2,2* gene is present in the B and D genomes but not in the A genome of hexaploid wheat.

Hybridization of Pro4 to hexaploid wheats along with tetraploids (*T. dicoccoides* and *T. turgidum*, AABB genomes) and the A-genome diploid *T. monococcum* (fig. 4*c*) suggested that an *Acc-2,4*–like gene was present in the A-genome diploid but was absent from the A genome in the tetraploids and hexaploids. From NT analysis of hexaploid wheat, it was known that Pro4 detected fragments only on chromosomes 3B and 3D. The tetraploid wheats, lacking the D genome, had only the chromosome 3B fragment. *Triticum monococcum* lacks both the B and the D genomes, but Pro4 hybridized to a fragment approximately the same size as the chromosome 3B fragment in the tetraploids and hexaploids. Presumably, this is a chromosome 3A fragment in *T. monococcum.* Furthermore, we tested three additional accessions each of the A-genome diploids, *T. monococcum* ssp. *monococcum* and *T. monococcum* ssp. *aegilopoides*, and the putative donor of the hexaploid A genome, *T. urartu.* All the accessions were collected from various regions of the Middle East. For each accession, Pro4 hybridized intensely to a single restriction fragment, indicating the presence of an *Acc-2,4*–like gene in the A genome. Based on phylogenetic analysis of the coding region described later in this paper, the *Acc-2,4* gene cloned from hexaploid wheat was assigned to the D genome.

It is probable that the A-genome *Acc-2,2*– and *Acc-2,4*–like genes are not found in hexaploid wheats because they were already absent from the tetraploid ancestor involved in the amphiploidization event that gave rise to hexaploid wheat. *Acc-2,4*, and probably *Acc-2,2*, was most likely eliminated from the A genome of tetraploids sometime after the formation of the ancestral tetraploid. This may reflect the evolutionary phenomenon of diploidization, for which it has been suggested that speciation through allopolyploidy may be accompanied by a rapid, nonrandom elimination of specific, low-copy DNA sequences at the early stages of allopolyploidization (Feldman et al. 1997). It is also possible that some functional copies of the gene escaped detection with gene-specific probes because parts of the gene containing target sites for the probes were deleted or diverged beyond the point of detection by hybridization.

No gene-specific probes for *Acc-2,3* were used in this study. The gene-specific probes for *Acc-2,1, Acc-2,2,* and *Acc-2,4* all detected the same restriction fragments, except that probes for *Acc-2,2* and *Acc-2,4* detected no fragment on chromosome 3A. This result and the number of different genes detected by sequencing suggest that multiple copies of the cytosolic ACCase gene are present in the wheat genomes on the long arms of group 3 chromosomes, possibly in tandem repeats.

The gene-specific probes were targeted to the least conserved sites in the gene to eliminate cross-hybridization. Sequence comparison of probe Pro1 with the avail-

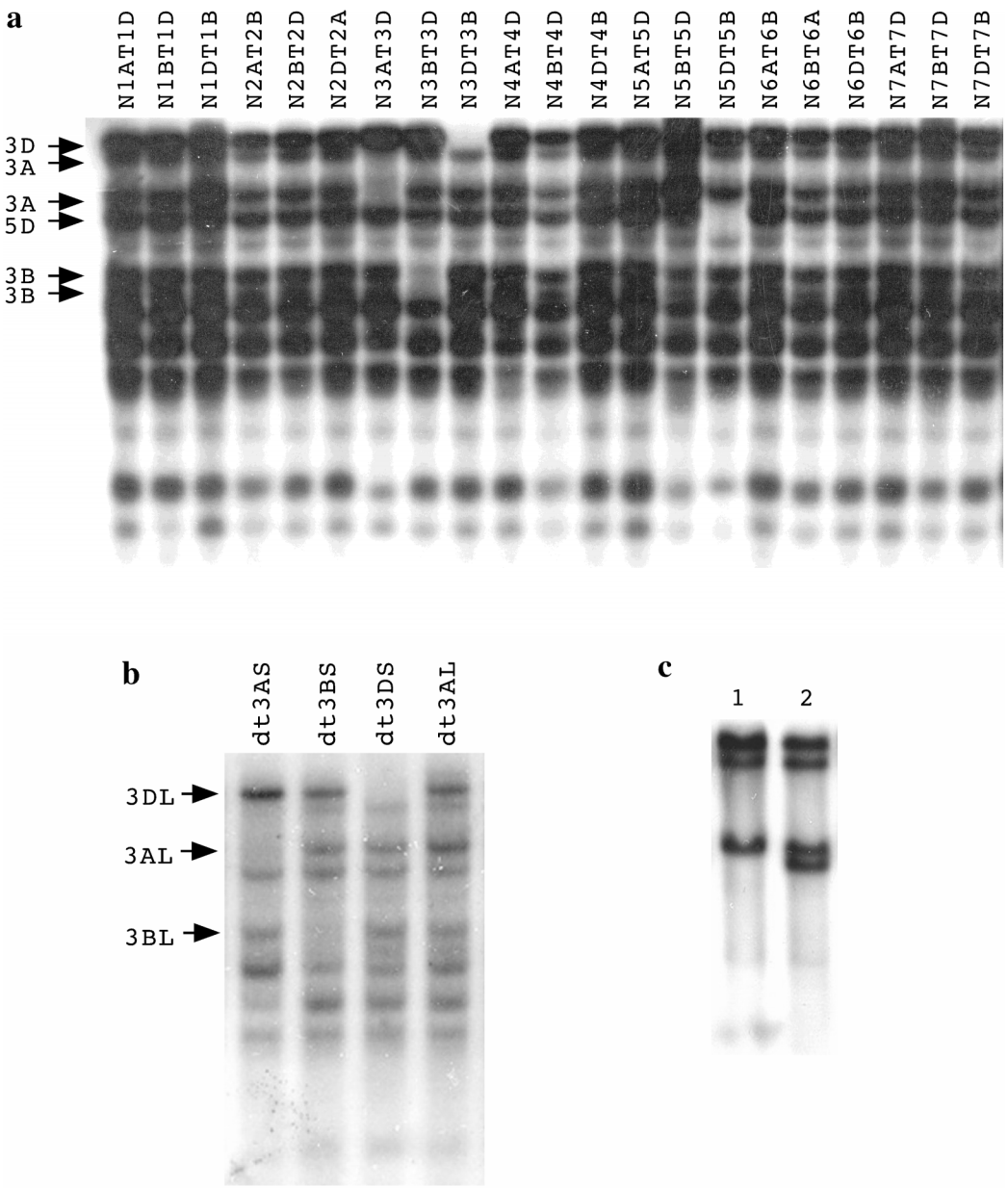

FIG. 2.—Chromosome mapping of *Acc-2* genes in wheat. *a,* Nullisomic-tetrasomic analysis with probe ucg22. Each lane is nullisomic for a different pair of chromosomes, and all 21 chromosomes are represented in the nullisomic condition. Fragments detected by ucg22 are absent in stocks nullisomic for chromosomes 3A, 3B, 3D, and 5D, indicating that the corresponding fragments map to these chromosomes and genomes. Genomic DNA was digested with *Hin*dIII. In a separate experiment, probe ucg21 revealed fragments on the same chromosomes. *b,* Lanes 1–3, dt lines missing the long arms of chromosomes 3A (dt3AS), 3B (dt3BS), and 3D (dt3DS), respectively. Lane 4, dt line missing the short arm of chromosome 3A (dt3AL). Missing bands in lanes 1–3 indicate that the probe detects the target gene on the long arm of chromosomes 3A, 3B, and 3D. Lane 4 shows all bands revealed by the probe in hexaploid wheat. In a similar experiment, the target gene was detected on the long arm of chromosome 5D. Genomic DNA was digested with *Hin*dIII and hybridized with probe ucg22. *c,* Lane 1, chromosome 3D deletion line with the fraction point at 0.78 on the long arm. Lane 2, euploid Chinese Spring wheat with the full chromosome complement. The missing band in lane 1 indicates that the probe detects the target gene in the distal 22% of the long arm of chromosome 3D. Each of the other bands seen in lane 1 was missing in lines carrying deletion of either chromosome 3A, chromosome 3B, or chromosome 5D. Genomic DNA was digested with *Eco*RV and hybridized with probe ucg22.

able sequence of gene *Acc-2,2,* and of Pro21 with the available sequence of gene *Acc-2,1* reveals no significant sequence similarity to allow cross-hybridization. Short stretches of identical sequence existed in Pro21 and *Acc-2,4* and in Pro4 and *Acc-2,2,* indicating the possibility of a weak signal from cross-hybridization. Such a weak signal, if scored as positive, could lead to a false estimate of gene copy number and assignment of individual genes to wrong genomes. This problem was overcome by using multiple gene-specific probes. All gene-specific probes except for Fl5′1 and Fl5′2 (fig. 1) lie within ACCase genes. Probes Fl5′1 and Fl5′2, which target the 5′ flanking regions of the genes, as well as probe Int2, which targets the first intron of the gene,

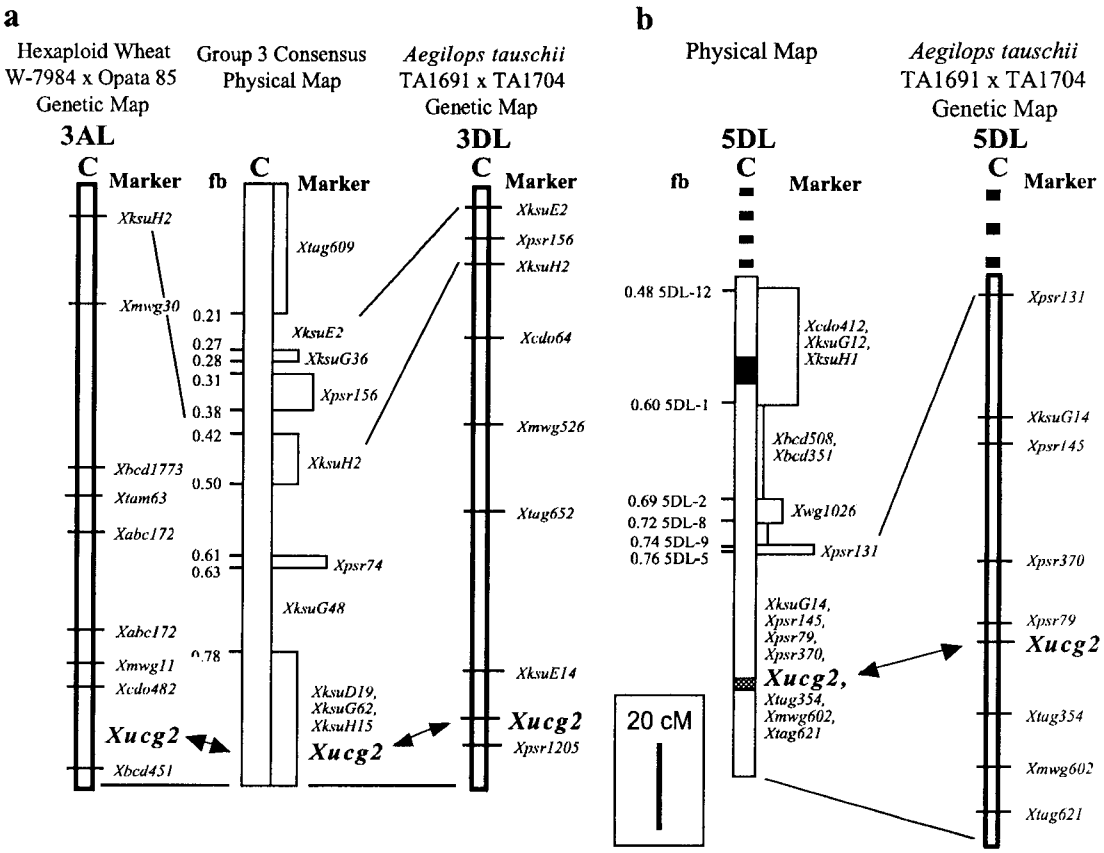

FIG. 3.—Cytosolic ACCase loci on wheat chromosomes. *a,* Genetic and physical maps of the long arms of group 3 chromosomes (Gill et al. 1991; Delaney et al. 1995*b*; Nelson et al. 1995). ''C'' marks the position of the centromere. Selected genetic markers are shown to the right of the genetic maps. Fraction breakpoints (fb) are indicated to the left of physical maps (e.g., fb 0.78 means 22% of the long arm is missing), and deletion intervals and selected markers mapping within them are indicated to the right. On the genetic map of the long arm of chromosome 3A of hexaploid wheat, marker *Xucg2* corresponds to probes ucg21, Pro1, and Int1. These probes all cosegregated and mapped with a LOD value of less than 2.0 to the indicated interval. On the consensus physical map, *Xucg2* mapped distal to fraction breakpoint 0.78. On the genetic map of the long arm of chromosome 3D of *Aegilops tauschii,* marker *Xucg2* corresponds to probes ucg21, Pro1, Pro4, Pro21, and Int1. These probes all cosegregated and mapped near the distal end of the chromosome arm. Arrows indicate positions of the *Acc-2* loci. The bar indicates the scale for the genetic maps in centimorgans (cM). *b,* Physical and genetic maps of the distal portion of the long arm of chromosome 5D of *A. tauschii* (Gill et al. 1996*a*). Map descriptions are the same as those for *a.* On the physical map, marker *Xucg2* mapped in the distal 24% of the long arm of chromosome 5D. On the genetic map, marker *Xucg2* mapped in the distal region of the long arm of *A. tauschii* chromosome 5D.

hybridized with multiple fragments, suggesting that they contain repetitive sequence elements. The 5′ end of *Acc-2,1* appears to contain a potentially multicopy retrotransposon-like element (fig. 1).

Unknown Character of the *Acc-2*–Related Sequences Detected on Chromosome 5D

None of the gene-specific probes hybridized to the locus on chromosome 5D identified with probes ucg21 and ucg22. The 5D locus could contain either an active gene or yet another pseudogene. A functional cytosolic ACCase gene located on chromosome 5D would have to be significantly different in its sequence in the untranslated regions to escape detection by the gene-specific probes used in this study. The additional copy of a cytosolic ACCase gene with divergent promoter regions and specific chromosome localization suggested a gene duplication/translocation event in the D genome lineage. Such a gene may already have acquired an alternative function. Identifying this function will require cloning and sequencing of a genomic fragment containing the *Acc-2*–related sequences from chromosome 5D.

Acc-2–Related Pseudogene

An *Acc-2* related pseudogene was identified by sequencing a 5-kb DNA fragment isolated from a hexaploid wheat genomic library (Podkowinski et al. 1996). Its sequence revealed a partially processed condition. The pseudogene (*Ψ-Acc-2,1*) contained only 2 out of the 13 introns expected based on the intron/exon structure of *Acc-2* genes and had a 7-nt (TTTCAAC) repeat, creating a frameshift that would prevent synthesis of a full-length ACCase (fig. 1), confirming the nonfunctional character of *Ψ-Acc-2.* A fragment of a second copy of the pseudogene (*Ψ-Acc-2,2*) was cloned by PCR. The two sequences differed by only 3% within a 350-bp fragment; however, the second copy did not have the 7-nt repeat. No other frameshifts or early stop codons were found in either copy of the pseudogene. The possibility of the pseudogene being located exclusively in the *Acc-*

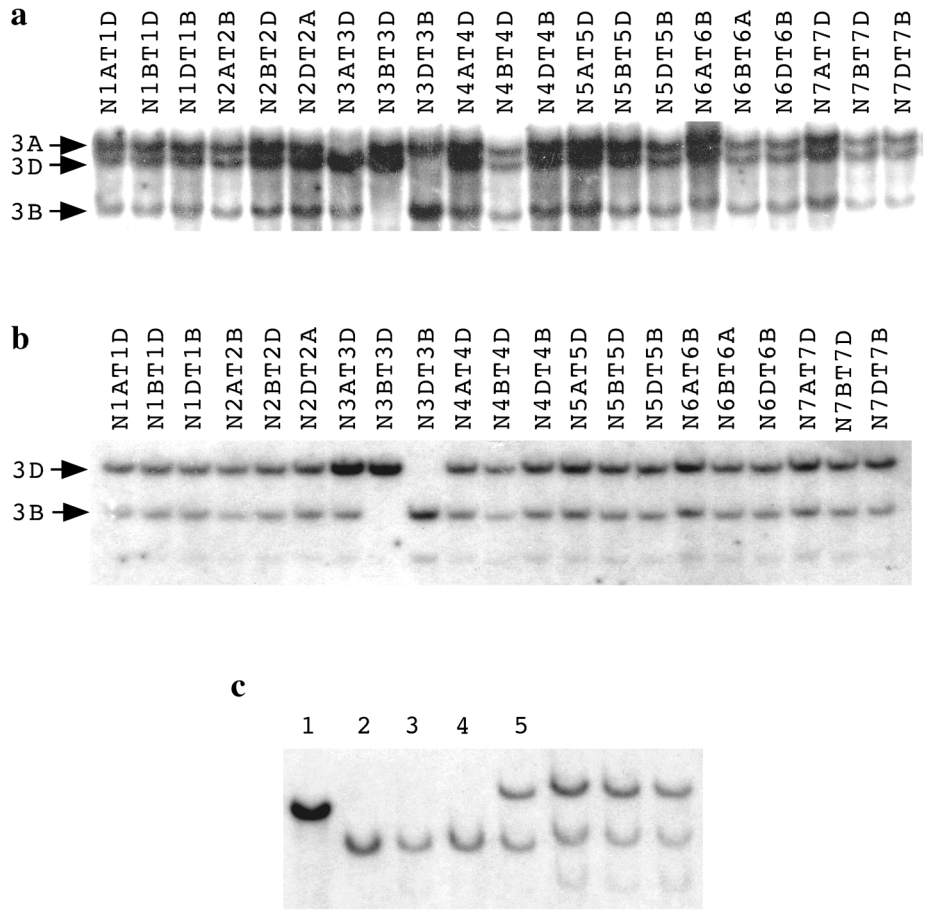

FIG. 4.—Chromosome mapping of *Acc-2* genes in wheat with gene-specific probes. *a,* Nullisomic-tetrasomic analysis with probe Int1 (gene *Acc-2,1*) identified a single restriction fragment on each of the group 3 chromosomes, 3A, 3B, and 3D, but none on chromosome 5D. Genomic DNA was digested with *Eco*RI. *b,* Nullisomic-tetrasomic analysis with probe Pro4 (gene *Acc-2,4*) identified a single restriction fragment on group 3 chromosomes 3B and 3D, but none on chromosomes 3A and 5D. Genomic DNA was digested with *Eco*RI. *c,* Southern hybridization with probe Pro4 (gene *Acc-2,4*). D-genome diploid *Aegilops tauschii* (lane 1), A-genome diploid *Triticum monococcum* (lane 2), wild AABB tetraploid *Triticum dicoccoides* (lane 3), cultivated AABB tetraploid *Triticum turgidum* (lane 4), and four *Triticum aestivum* AABBDD hexaploids, W-7984, Opata 85, Chinese Spring, and TAM107 (lanes 5, 6, 7, and 8, respectively). Genomic DNA was digested with *Eco*RI.

2 locus on chromosome 5D was excluded by the PCR experiment. A fragment of the pseudogene was amplified when genomic DNA from NT lines missing chromosome 5D was used as template. The identity of the DNA fragment was verified by sequencing cloned PCR products. The pseudogene was found in other Triticeae species (unpublished results). A copy of the pseudogene from the D-genome diploid *A. tauschii* ssp. *meyeri* (accession number 1691) was identical to *Ψ-Acc-2,1* from hexaploid wheat. A copy of the pseudogene from the A-genome diploid *T. urartu* (accession number 763) was similar to another pseudogene copy from hexaploid wheat (not shown). Therefore, we postulate that in hexaploid wheat, a copy of the pseudogene is located in the *Xucg2* locus on each of the homoeologous group 3 chromosomes (A, B, and D). The intronless structure of the pseudogene can be explained by integration of a cDNA fragment formed by reverse transcription into one of the gene copies present in the *Acc-2* locus. This scenario is consistent with the localization of the pseudogene in this locus and the presence of some introns at one end of the pseudogene. These remaining introns came from the active gene which was a target of the integration event.

Intron Loss

Intron loss or gain is another event detected in *Acc* genes. Two introns are missing from the entire cytosolic clade (Podkowinski et al. 1996; Gornicki et al. 1997). One barley *Acc-2* gene (cyt2) has still two introns fewer in the gene region used for the phylogenetic analysis. Creation of partially processed pseudogenes with many fewer introns is another type of intron reduction event that adds complexity to the *Acc-2* loci by creating nonfunctional genes.

Phylogenetic Analysis of *Acc-2* Genes of the Triticeae Tribe

A segment of the ACCase gene (fig. 1) with its ends anchored in highly conserved exons encoding the biotin carboxylase domain, which itself is the most conserved domain of ACCase, was selected for phyloge-

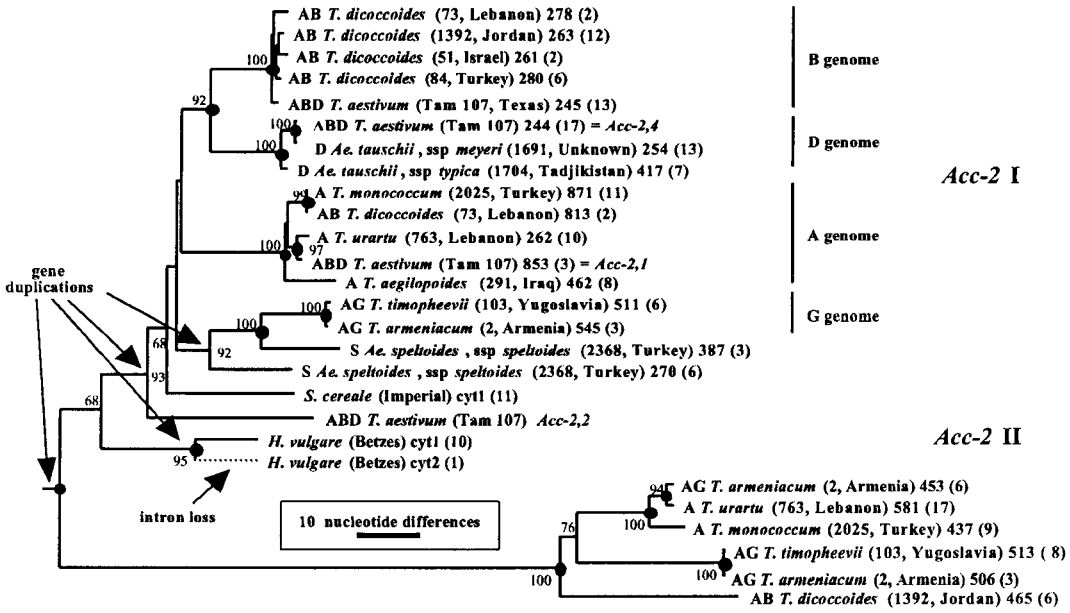

FIG. 5.—Phylogenetic analysis of *Acc-2* genes of the Triticeae tribe. A phylogenetic tree was created by the Neighbor-Joining method based on the alignment of merged intron sequences with distances not corrected for multiple substitutions and gaps excluded from pairwise comparisons. The alignment was 1,385 nt long and contained 855 invariant positions. Taxon names include genome composition (for *Triticum* and *Aegilops* species), species name, accession number/cultivar name, geographical origin, sequence ID number, and number of clones analyzed. The two major clades are named *Acc-2*I and *Acc-2*II. Sequence of *Lolium rigidum* (AUS92, Australia) cyt1 (1) (GenBank AF343454) introns was used as the outgroup. Bootstrap values are shown at selected nodes as percentages of 1,000 replicates. Nodes with strong statistical support found on trees generated by different methods are indicated with dots. Gene sequences were deposited in GenBank with accession numbers AF306803–AF306829.

netic analysis of *Acc* genes because of universal applicability in Poaceae and because homology assessment of the nucleotide or amino acid positions was straightforward. Therefore, phylogenetic analysis can include the multidomain cytosolic and plastid ACCases from all plants. *Acc-2*–specific primers, based on the available sequence information for plant ACCases, were relatively long (24-mer), and their target sites were offset relative to each other by several nucleotides. Various combinations of these primers worked effectively in PCR amplification at low annealing temperature even with some mismatches. Primer pairs yielding fewer nonspecific products were tested on multiple species from the *Triticum*/*Aegilops* complex, as well as on barley, rye, and *Lolium* (*L. rigidum*), for their ability to amplify fragments of multiple copies of the cytosolic ACCase gene, as well as the related pseudogene. These primers did not amplify the corresponding fragment of the plastid ACCase gene. The PCR-cloned *Acc-2* gene fragment was about 1.8 kb long and included approximately the same number of nucleotides in six exons and five introns. The corresponding fragment of the pseudogene was about 0.7 kb long. Plant species included in the analysis are identified in figure 5. All positions of the intron part of this alignment were used to calculate the neighbor-joining tree shown in figure 5.

Including intron sequences of the *Acc-2* genes was essential for the phylogenetic analysis below the tribe level, where most nucleotide changes were found in introns. The resolution remained sufficient at the level of the major genomes of the *Triticum*/*Aegilops* complex but declined quickly when sequences of homologous genomes were compared, e.g., the D genome of *T. aestivum* and *A. tauschii*, or the A genome of *T. aestivum*, *T. urartu*, *T. monococcum*, and *T. dicoccoides* (fig. 5). Creating multiple alignments of sequences belonging to clade *Acc-2*I or *Acc-2*II posed no problem. Creating alignments of intron sequence became more difficult when the two major clades (*Acc-2*I and *Acc-2*II) and an outgroup were included in the analysis. However, small changes in the intron sequence alignment introduced by hand did not alter the topology of the major clades of the tree, as the alignment of these subsets of sequences was not altered, and introduced only minor changes in the genetic distances between clades *Acc-2*I and *Acc-2*II. *Lolium Acc-2* sequence was used as an outgroup only. Its alignment with the remaining sequences was problematic at several intron sites. Introns became too divergent to allow multiple alignments for phylogenetic analysis at or above the tribe level. This was also true for genes encoding cytosolic and plastid multidomain ACCases in Poaceae. Exon sequences are suitable for phylogenetic analysis of ACCase genes at this level of genetic relatedness (unpublished results).

How Many Copies of the *Acc-2* Gene Are Present in Wheat Genomes?

The results of sequence analysis of genomic DNA and cDNA obtained from various cloning experiments suggest that there are more than three copies of the *Acc-2* gene in hexaploid wheat (Gornicki et al. 1994; Podkowinski et al. 1996). Fragments of five genes, *Acc-2,1*–*Acc-2,5*, were isolated from genomic library or cloned

by PCR targeting the 5′ end of the gene. Four copies of the *Acc-2* gene from hexaploid wheat were included in the phylogenetic analysis described above: two sequences assigned to previously identified genes *Acc-2,1* and *Acc-2,4* by virtue of their identity with sequence of overlapping genomic and corresponding cDNA fragments, one genomic sequence which could not be assigned to any of the other genes mentioned above, and a sequence of gene *Acc-2,2* which was cloned from a genomic library. Genes *Acc-2,3* and *Acc-2,5* could not be included in the phylogenetic analysis because the cloned fragments corresponded to a different part of the gene. Two copies of the *Acc-2* gene were found in several diploid Triticeae species: *T. urartu*, *T. monococcum*, *A. speltoides*, and barley (fig. 5). Three *Acc-2* genes were found in diploid *L. rigidum* (GenBank accession numbers AF343454–AF343456).

Identification of all copies of the *Acc-2* gene family may prove difficult. On the one hand, the PCR-cloning experiment is expected to fail to detect all those copies of the gene that accumulated multiple mutations within the primer target sites. For example, primers designed for the amplification of the known cytosolic ACCase genes did not work for the plastid ACCase genes, although the latter were created by duplication of the former. The *Acc-2*–related gene(s) identified by mapping experiments on chromosome 5D may represent another such case. This deficiency could probably be overcome by additional PCR-cloning experiments with new primers designed based on the much larger sequence database which is now available. Analysis of a large number of clones should overcome any bias in PCR amplification. On the other hand, recent gene duplication events that have not had enough time for mutations to accumulate would escape detection.

The phylogenetic analysis suggests that gene duplication was an important recurring event shaping the *Acc-2* gene family. One duplication event led to the creation of the two major clades (*Acc-2*I and *Acc-2*II; fig. 5). It occurred before the divergence of the Triticeae species. The A genomes of diploids *T. urartu* and *T. monococcum*, and probably the A genome of the AG tetraploid *T. armeniacum*, have both copies of the gene. The topology of clade *Acc-2*II suggests that this copy of the gene is present in genomes B and G as well, although the A genome assignment of all the genes in clade *Acc-2*II cannot be strictly ruled out. It is then surprising that an *Acc-2*II clade gene was not found in the hexaploid wheat for which a close relative of *T. urartu* donated the A genome via an AABB tetraploid. The D genome was also not represented in clade *Acc-2*II.

It is possible that both copies of the gene, *Acc-2*I and *Acc-2*II, are retained only is some genomes: the A genome of A-genome diploids and AB-genome tetraploids, the B genome of AB-genome tetraploids, and the G genome of AG-genome tetraploids. The *Acc-2*II copy was probably lost from other diploid species and independently from the A genome of hexaploid wheat. The A-genome *Acc-2*I copy of the gene was not detected in AG tetraploids. Therefore, it is possible that this copy was lost from these species and, instead, the *Acc-2*II copy was retained in their A genomes (fig. 5). Some of these apparent gene loss events could be explained by a failure of our PCR-based experiment to amplify all existing gene copies despite a significant number of clones having been analyzed.

The phylogenetic analysis further suggests that independent *Acc-2* gene duplication events occurred in some other lineages (fig. 5). Furthermore, none of the genes for which gene-specific probes were available mapped on chromosome 5D. The presence of the *Acc-2*II gene in A-genome diploids and the absence of this gene in D-genome diploids provides an argument against the possibility that the *Acc-2* gene of chromosome 5D belongs to the *Acc-2*II clade. Therefore, it is likely that the chromosome 5D locus contains an additional copy or additional copies of the gene.

Assessment of Substitution Rate Variation in Different Segments of the *Acc-2* Genes

Nucleotide substitution rates in introns and at synonymous positions among *Acc-2*I genes belonging to clades A, B, and D were compared for the gene fragment used in phylogenetic analysis (fig. 6*a*). For all three pairwise comparisons, *Acc-2*IA and *Acc-2*IB, *Acc-2*IA and *Acc-2*ID, and *Acc-2*IB and *Acc-2*ID, substitution rates in introns were similar to each other, and synonymous substitution rates were approximately twice as high. Little variation was observed when rates for pairs of putative orthologs were compared (reflected in low standard deviations). Intron 6 in genes belonging to clade *Acc-2*IB is an exception. For an unknown reason, this intron is much more similar to the corresponding intron in *Acc-2*ID genes than the remaining four introns. This is reflected in the high standard deviation (introns 5–9, average for the *Acc-2*IB : *Acc-2*ID rate; fig. 6*a*) and, because of the length of this intron, in an apparently shorter distance between taxa of clades B and D (fig. 5). Synonymous substitution rates vary significantly from exon to exon, but most of this variability results from too few substitutions being counted in the shortest exons.

Nucleotide substitution rates in the 5′-end noncoding portion of *Acc-2*, including the first exon and the first intron (the translation start codon is located in the second exon), were assessed for three genes for which such genomic sequences are available (fig. 6*b*). Substitution rates in different segments of the 5′-end noncoding part of genes *Acc-2,1* and *Acc-2,4*, putative orthologs located on chromosomes 3A and 3D, respectively, were compared with average intron rates between genes belonging to clades *Acc-2*IA and *Acc-2*ID. The rate in the promoter region (block Y), the leader, and the first intron are approximately two times as high as the rates for the coding-region introns and similar to or slightly higher than the synonymous rates (fig. 6*a*). The nucleotide substitution rate for block X, which is separated from block Y by a nonconserved expansion element (fig. 1), is dramatically higher than the rates for the rest of the gene. These results suggest an accelerated mutation rate at the 5′ end of the gene. Precise identification of the 5′ end

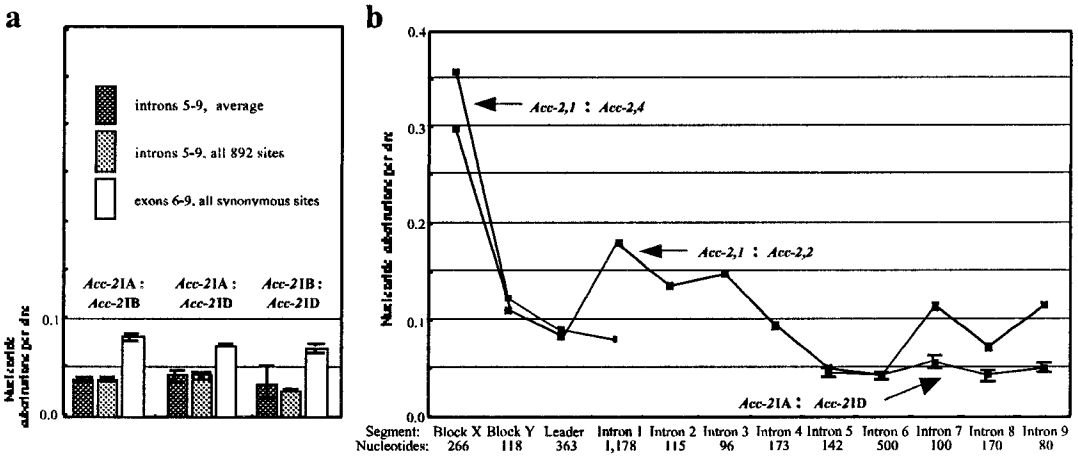

FIG. 6.—Nucleotide substitution rates in different parts of the *Acc-2* gene. *a,* Average substitution rates in introns 5–9 and exons 6–9 of genes assigned to clades *Acc-2*IA, *Acc-2*IB, and *Acc-2*ID are shown with standard deviations. Intron rates were calculated in two different ways. First, average rates were calculated for each intron separately, and then those rates were averaged (''introns 5–9, average''). Second, average rates were calculated for all 892 intron sites as one set based on merged alignment of intron sequences (''introns 5–9, all 892 sites''). Gaps were excluded only from pairwise comparisons. Synonymous substitution rates were calculated based on alignment of merged exon sequences (''exons 6–9, all synonymous sites''). *b,* Substitution rates in different segments of the 5′-end portion of *Acc-2,1* and *Acc-2,4* genes (''*Acc-2,1* : *Acc-2,4*'') are compared with average substitution rates in introns 5–9 of all available genes assigned to clades *Acc-2*IA and *Acc-2*ID (''*Acc-2*IA : *Acc-2*ID''). Genes *Acc-2,1* and *Acc-2,4* belong to these two clades, respectively. Substitution rates in different segments of *Acc2,1* and *Acc-2,2* genes (''*Acc-2,1* : *Acc-2,2*'') are also shown. Gaps were excluded only from pairwise comparisons, and standard deviations are shown for the average rates. The size of each segment is shown in nucleotides. The structure of *Acc-2* genes is shown in figure 1.

of the gene is not possible, but conservation of two blocks of sequences at the 5′ end hints at their possible significance as part of the gene. A similar comparison for a pair of putative paralogs, *Acc-2,1* and *Acc-2,2,* confirmed the high substitution rate in block X but also revealed high variability of the nucleotide substitution rate for different introns.

These results point to potential problems that may be encountered in this type of phylogenetic analysis due to substitution rate variability between different gene segments leading to incongruent phylogenies. For example, nucleotide substitution rates between genes *Acc-2,1* and *Acc-2,2,* and between clades *Acc-2*IB and *Acc-2*ID calculated based on intron sequences can vary significantly from intron to intron. Comparison of genes with different numbers of introns, such as barley genes cyt1 and cyt2 (fig. 5), could further skew the outcome of the analysis. Our system allows detection and evaluation of such phenomena. The final conclusions are based on sequence comparisons of multiple introns and exons.

Phylogenetic Inferences Based on *Acc-2* Genes

The complex pattern of evolution, coupled with the limited set of sequences used in the analysis, made establishing orthology difficult. The fact that the genes seemed to form major clades consistent with their putative chromosome assignments (A, B, D, G) or taxonomic relationships (e.g., between Hordeum and Triticum) can only be taken as an indication of possible orthologous relationships.

Many of the branch points on the Neighbor-Joining tree shown in figure 5 were found with equally good statistical support on other neighbor-joining trees and on consensus maximum-parsimony trees. These well-supported clades reflect correctly known relationships among *Triticum* species (Cox 1998), supporting orthologous relationships between these groups of genes. For example, the origins of A, B, and D genomes in hexaploid wheat can be traced to *T. urartu,* AB tetraploids, and *A. tauschii,* respectively. A, B, D, and G genomes diverged early in the evolution of the *Triticum/Aegilops* complex, whereas the polyploid species had a much more recent origin. Some of these relationships can be observed for both major *Acc-2* clades (I and II).

The sequence of some other evolutionary events is ambiguous. This ambiguity is, at least in part, due to insufficient resolution of the method. It appears that the diploid progenitors of the Triticum species (donors of A, B, D, and G genomes) radiated at approximately the same time. Missing genes, lost from some genomes or simply not cloned in our experiments, could lead to paralogs being mistaken for orthologs. Finally, it is conceivable that a copy of a multicopy gene family is a pseudogene. This could be difficult to determine without sequencing the entire gene and may require analysis of gene expression. No such problems are evident in our phylogenetic analysis, but they could lead to erroneous phylogenetic inferences.

Discussion

The goal of our analysis was to provide information on the evolutionary events and processes that shape specific loci containing genes encoding cytosolic ACCase, a key enzyme of very–long-chain fatty acid biosynthesis. Good understanding of the structure and history of the gene family is an essential component of functional analysis of *Acc-2* genes and their products aimed at understanding the key steps of wheat metabolism, e.g., by providing information on gene copy number and pos-

sible divergence of gene function. The functional study, in turn, provides important information for the phylogenetic analysis, e.g., by identifying functional genes and assigning specific functions to individual genes.

Our approach combined cytogenetic analysis with a molecular evolutionary study of this small gene family. First, multiple cytosolic ACCase genes from wheat were analyzed to reveal their structures. Second, the *Acc-2* loci were mapped on wheat chromosomes. Finally, analysis of *Acc-2* genes in wheat species of different ploidys using Southern hybridization, cloning, and sequencing enabled bridging of the gap between the cytogenetic and molecular approach by assigning gene sequences to specific loci on homoeologous chromosomes of known genomic affinity on the one hand, and by reconstructing phylogenetic relationships among them on the other hand. Nucleotide substitution rates in different segments of the gene were assessed.

Cytosolic ACCase is encoded by a small gene family with more than one copy of the gene present in at least some of the Triticeae genomes. In the reconstruction of the evolution of the *Acc-2* gene, we documented gene duplication events that occurred in the polyploid species and/or their diploid ancestors and postulated that these duplications were followed by deletion or loss of function of some gene copies in different genomes and plant lineages. Gene loss due to deletion, homogenization, or formation of pseudogenes followed by accelerated mutation that occurred in parallel with gene duplication seems possible. All of these processes are likely to be more frequent in polyploid species. The end result of such recent evolutionary events would be a different set of active genes even in closely related species.

Evolutionary changes at the 5′ end of the gene including promoter/regulatory elements, leader sequences, first exons, and introns are of special interest, as they may lead to important changes in gene function. The structure of promoter and regulatory elements and the activity of individual *Acc* genes and their tissue and development specificities are being investigated (unpublished results). These results indicate that multiple *Acc-2* genes, including genes *Acc-2,1*, *Acc-2,2,* and *Acc-2,4*, discussed in this paper, are transcriptionally active and show distinct expression patterns in wheat plants. Evolution of the 5′-end region of the gene may be affected by events in the neighboring intergenic region, e.g., by insertion of transposable elements that occasionally invade the gene itself, affecting gene function. Higher substitution rates and the presence of a highly variable expansion elements may be signatures of such events. A better understanding of the evolution of the *Acc-2* gene family makes a study of such border phenomena feasible.

Recurring duplication of the *Acc* gene in different plant lineages has played an important role in providing genetic material for the creation of new genes, such as the plastid ACCase gene, which occurred more then once during plant evolution: in Poaceae (Konishi et al. 1996), in Brassicaceae (Schulte et al. 1997), and in Geraniaceae (Christopher and Holtum 2000). This duplication event occurred at the onset of grass family evolution and was followed by a chromosome translocation. Additional copies of the *Acc-2* gene that were created later during grass evolution may have or may eventually acquire a different function, e.g., different tissue or development specificity. Such duplication events, followed in some instances by translocation to a new locus on a different chromosome and/or potential deletion or inactivation of some gene copies accompanied by variable substitution rates in different gene parts, different genes, and different chromosome loci, add plasticity and shape the grass genome. These are probably very common locus/gene-specific events occurring in the context of large chromosomal rearrangements, frequent polyploidization, and rapid changes in the intergenic regions.

Supplementary Material

GenBank accession numbers are as follows: *Acc-2,1*, AF305204; *Acc-2,2*, AF305205; *Acc-2,3*, AF305206; *Acc-2,4*, AF305207; *Acc-2,5*, AF305208; *Ψ-Acc-2,1*, AF305209; *Ψ-Acc-2,2*, AF362956, *Acc-2* sequences from *Triticum* and *Aegilops* species, AF306803–AF306829.

Acknowledgments

We thank W. J. Buikema for his help with the sequencing part of the project, and J. Wendel and J. Jelenska for comments on the manuscript. DNA sequencing was performed by the University of Chicago Cancer Center DNA Sequencing Facility. This work was supported by a gift from Monsanto Corp.

LITERATURE CITED

BENNETZEN, J. L. 2000. Transposable element contribution to plant gene and genome evolution. Plant Mol. Biol. **42**:251–269.

BENNETZEN, J. L., P. SANMIGUEL, M. CHEN, A. TIKHONOV, M. FRANCKI, and Z. AVRAMOVA. 1998. Grass genomes. Proc. Natl. Acad. Sci. USA **95**:1975–1978.

CHRISTOPHER, J. T., and J. A. M. HOLTUM. 2000. Dicotyledons lacking the multisubunit form of the herbicide-target enzyme acetyl coenzyme A carboxylase may be restricted to the family Geraniaceae. Aust. J. Plant Physiol. **27**:845–850.

COX, T. S. 1998. Deepening the wheat gene pool. J. Crop Production **1**:1–25.

DELANEY, D., S. NASUDA, T. R. ENDO, B. S. GILL, and S. H. HULBERT. 1995*a*. Cytogenetically based physical maps of the group-2 chromosomes of wheat. Theor. Appl. Genet. **91**: 568–573.

———. 1995*b*. Cytogenetically based physical maps of the group-3 chromosomes of wheat. Theor. Appl. Genet. **91**: 780–782.

DVORAK, J., P. DITERLIZZI, H.-B. ZHANG, and P. RESTA. 1993. The evolution of polyploid wheats: identification of the A genome donor species. Genome **36**:21–31.

DVORAK, J., M.-C. LUO, Z.-L. YANG, and H.-B. ZHANG. 1998. The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor. Appl. Genet. **97**:657–670.

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

EGLI, M., S. LUTZ, D. SOMERS, and B. GENGENBACH. 1995. A maize acetyl-coenzyme A carboxylase cDNA sequence. Plant Physiol. **108**:1299–1300.

ENDO, T. R., and B. S. GILL. 1996. The deletion stocks of common wheat. J. Hered. **87**:295–307.

FARIS, J. D., K. M. HAEN, and B. S. GILL. 2000. Saturation mapping of a gene-rich recombination hot spot region in wheat. Genetics **154**:823–835.

FEDOROFF, N. 2000. Transposons and genome evolution in plants. Proc. Natl. Acad. Sci. USA **97**:7002–7007.

FELDMAN, M., B. LIU, G. SEGAL, S. ABBO, A. LEVY, and J. VEGA. 1997. Rapid elimination of low-copy DNA sequences in polyploid wheat: a possible mechanism for differentiation of homoeologous chromosomes. Genetics **147**:1381–1387.

GALE, M. D., and K. M. DEVOS. 1998. Comparative genetics in the grasses. Proc. Natl. Acad. Sci. USA **95**:1971–1974.

GAUT, B. S. 1998. Molecular clocks and nucleotide substitution rates in higher plants. Evol. Biol. **30**:93–120.

GILL, K. S., B. S. GILL, and T. R. ENDO. 1993. A chromosome region-specific mapping strategy reveals gene-rich telomeric ends in wheat. Chromosoma **102**:374–381.

GILL, K. S., B. S. GILL, T. R. ENDO, and E. V. BOIKO. 1996*a*. Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics **143**: 1001–1012.

GILL, K. S., B. S. GILL, T. R. ENDO, and T. TAYLOR. 1996*b*. Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics **144**: 1883–1891.

GILL, K. S., E. L. LUBBERS, B. S. GILL, W. J. RAUPP, and T. S. COX. 1991. A genetic linkage map of *Triticum tauschii* (DD) and its relationship to the D genome of bread wheat (AABBDD). Genome **34**:362–374.

GORNICKI, P., J. FARIS, I. KING, J. PODKOWINSKI, B. GILL, and R. HASELKORN. 1997. Plastid-localized acetyl-CoA carboxylase of bread wheat is encoded by a single gene on each of the three ancestral chromosome sets. Proc. Natl. Acad. Sci. USA **94**:14179–14185.

GORNICKI, P., and R. HASELKORN. 1993. Wheat acetyl-CoA carboxylase. Plant Mol. Biol. **22**:547–552.

GORNICKI, P., J. PODKOWINSKI, L. A. SCAPPINO, J. DIMAIO, E. WARD, and R. HASELKORN. 1994. Wheat acetyl-CoA carboxylase: cDNA and protein structure. Proc. Natl. Acad. Sci. USA **91**:6860–6864.

HOHMANN, U., T. R. ENDO, K. S. GILL, and B. S. GILL. 1994. Comparison of genetic and physical maps of group 7 chromosomes from *Triticum aestivum* L. Mol. Gen. Genet. **245**: 644–653.

INCLEDON, B. J., and J. C. HALL. 1997. Acetyl-coenzyme A carboxylase: quaternary structure and inhibition by graminicidal herbicides. Pestic. Biochem. Physiol. **57**:255–271.

JOACHIMIAK, M., G. TEVZADZE, J. PODKOWINSKI, R. HASELKORN, and P. GORNICKI. 1997. Wheat cytosolic acetyl-CoA carboxylase complements an *ACC1* null mutation in yeast. Proc. Natl. Acad. Sci. USA **94**:9990–9995.

KAM-MORGAN, L. N. W., B. S. GILL, and S. MUTHUKRISHNAN. 1989. DNA restriction fragment length polymorphisms: a strategy for genetic mapping of D genome of wheat. Genome **32**:724–732.

KIHARA, H. 1944. Discovery of the DD-analyser, one of the ancestors of vulgare wheats. Ag. Hort. (Tokyo) **19**:889–890.

KONISHI, T., K. SHINOHARA, K. YAMADA, and Y. SASAKI. 1996. Acetyl-CoA carboxylase in higher plants: most plants other than graminae have both the prokaryotic and the eukaryotic forms of this enzyme. Plant Cell Physiol. **37**:117–122.

KOSAMBI, D. D. 1944. The estimation of map distances from recombination values. Ann. Eugen. **12**:172–175.

LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY, S. E. LINCOLN, and L. NEWBURG. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics **1**:174–181.

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

MICKELSON-YOUNG, L., T. R. ENDO, and B. S. GILL. 1995. A cytogenetic ladder-map of wheat homoeologous group-4 chromosomes. Theor. Appl. Genet. **90**:1007–1011.

NELSON, J. C., A. E. VAN DEYNZE, E. AUTRIQUE, M. E. SORRELLS, Y. H. LU, S. NEGRE, M. BERNARD, and P. LEROY. 1995. Molecular mapping of wheat. Homoeologous group 3. Genome **38**:525–533.

PODKOWINSKI, J., G. E. SROGA, R. HASELKORN, and P. GORNICKI. 1996. Structure of a gene encoding a cytosolic acetyl-CoA carboxylase of hexaploid wheat. Proc. Natl. Acad. Sci. USA **93**:1870–1874.

SCHULTE, W., R. TOPFER, R. STRACKE, J. SCHELL, and N. MARTINI. 1997. Multi-functional acetyl-CoA carboxylase from *Brassica napus* is encoded by a multi-gene family: indication for plastidic localization of at least one isoform. Proc. Natl. Acad. Sci. USA **94**:3465–3470.

SEARS, E. R. 1944. The aneuploids of common wheat. Mo. Agr. Exp. Sta. Res. Bull. **572**:1–58.

———. 1966. Nullisomic-tetrasomic combinations in hexaploid wheat. Pp. 29–45 *in* R. RILEY and K. R. LEWIS, eds. Chromosome manipulation and plant genetics. Oliver and Boyd, Edinburgh, Scotland.

SEARS, E. R., and L. M. S. SEARS. 1978. The telocentric chromosomes of common wheat. Paper presented at Fifth International Wheat Genetics Symposium. Indian Society of Genetics and Plant Breeding, New Delhi, India.

TAKUMI, S., S. NASUDA, Y. G. LIU, and K. TSUNEWAKI. 1993. Wheat phylogeny determined by RFLP analysis of nuclear DNA. I. Einkorn wheat. Jpn. J. Genet. **68**:73–79.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.

UPADHYA, M. D., and M. S. SWAMINATHAN. 1963. Genome analysis in *Triticum zhukovskyi*, a new hexaploid wheat. Chromosoma **14**:589–600.

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

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

ELIZABETH KELLOGG, reviewing editor

Accepted May 14, 2001