© Springer-Verlag 2001

O.D. Anderson · C.C. Hsia · V. Torres

The wheat γ-gliadin genes: characterization of ten new sequences and further understanding of γ-gliadin gene family structure

Received: 24 August 2000 / Accepted: 15 December 2000

Abstract Ten new wheat γ-gliadin gene sequences are reported and an analysis of γ-gliadin gene family structure is carried out using all known γ-gliadin sequences. The new sequences comprise four genomic clones with significantly more flanking DNA than previously reported, and six cDNA clones from a wheat endosperm EST project. Analysis of extended flanking DNA from the genomic clones indicates the limits of conservation of γ-gliadin DNA sequence that are similar to those previously found with other gliadin and glutenin genes and that are theorized to define the DNA sequence necessary for gene control. Most of the flanking DNA is not homologous to any reported DNA sequence, and one flanking region contains the first MITE-like (miniature inverted transposable element) DNA sequence associated with gliadin genes. About a quarter of the encoded polypeptides would contain a free cysteine residue – an observation that may relate to reports that at least some gliadins can participate in wheat endosperm glutenin polymer formation. The new sequences represent both genes closely related to those previously reported and a new sub-class of γ-gliadins.

Keywords Gamma-gliadin · Gluten · Wheat · Quality · Evolution

Introduction

The major groups of wheat prolamins are the high-molecular-weight (HMW) glutenins encoded by genes on the long arm of the group-1 homoeologous chromosomes, the α-gliadins encoded on the group-6 chromosomes, the γ- and ω-gliadins, and the low-molecular-weight (LMW) glutenins all encoded by genes located on the short arm of the group-1 homoeologous chromosomes (Payne 1987; Singh and Shepherd 1988). The gliadin family is related to a number of other monocot storage proteins as well as those from more-distantly related dicot genera (Kreis et al. 1985; Shewry 1995).

DNA sequences have been reported for several γ-gliadin genes, including seven complete coding sequences from genomic and polymerase chain reaction (PCR)-derived clones (Rafalski 1986; Sugiyama et al. 1986; Scheets and Hedgcoth 1988; D'Ovidio et al. 1995; von Büren et al. 2000), two partial PCR-generated sequences (D'Ovidio et al. 1991; Maruyama et al. 1998), one complete cDNA sequence (Bartels et al. 1986), and one partial cDNA sequence (Scheets et al. 1985). These have come from a variety of sources: eight different cultivars representing bread, spelt, and durum wheats. We have been carrying out a detailed analysis of the wheat prolamin gene families using as a model the hard red winter wheat cultivar Cheyenne (HMW-glutenins, Anderson et al. 1988; α-gliadins, Anderson and Greene 1997; Anderson et al. 1997; LMW-glutenins, Cassidy et al. 1998; ω-gliadins, Hsia and Anderson 2001). No γ-gliadins are yet reported from this cultivar. To complete the analysis of Cheyenne prolamins we now report characterization of ten new γ-gliadin sequences.

Material and methods

A wheat cv Cheyenne genomic lambda bacteriophage (λ) library λSep6-CNN (Anderson et al. 1997) was screened using a 60-bp oligonucleotide (CAACAATGCTGCCAACAACTAGCACAGAT TCCTCAGCAGCTCCAGTGTGCAGCCATCCAT) annealed with a 15-bp 3 complementary primer (ATGGATGGCTGCACA) and extended in the presence of ^{32}P-dCTP and Sequenase DNA polymerase (US Biochemical). The 60-bp oligonucleotide sequence was taken from the non-repetitive main cysteine-containing coding domain of γ-gliadin clone W10 (Scheets et al. 1985) and included the sequence encoding γ-gliadin conserved cysteines C_{f1},

Communicated by J. Dvorak

O.D. Anderson (✉) · C.C. Hsia
U.S. Department of Agriculture, Agricultural Research Service,
Western Regional Research Center, 800 Buchanan Street,
Albany, CA 94710, USA
e-mail: oandersn@pw.usda.gov
Tel.: +510-559-5773, Fax: +510-559-5777

V. Torres
Instituto Nacional Investigaciones Agrarias, Madrid, Spain

Table 1 Wheat γ-gliadin sequences

Name	Sequence type[a]	Cultivar	Length DNA sequence	Access. no.[b]	References
W10	C[c]	Newton	798	M16060	Scheets et al. 1985
Tag1436	C	Chinese Spring	1,142	–[d]	Bartels et al. 1986
L311A	G	Yamhill	2,450	M13712	Rafalski 1986
L311B	G	Yamhill	2,450	M13713	Rafalski 1986
W1621	G	Yamhill	1,397	M16064	Sugiyama et al. 1986
W1020	G	Yamhill	2,086	M36999	Scheets and Hedgcoth 1988
TDA16	P[c]	Langdon[e]	850	X53412	D'Ovidio et al. 1991
TD9	P[c]	Lira	909	X77963	D'Ovidio et al. 1995
EWγ	P[c]	1CW	840	D78183	Maruyama et al. 1998
GAG56dok	P	Oberkulmer[f]	947	AF120267	von Büren et al. 2000
GAG56dfo	P	Forno	956	AF144104	von Büren et al. 2000
γ13	G	Cheyenne	1,843	AF234646	This paper
G1	G	Cheyenne	5,718	AF234647	This paper
G6	G	Cheyenne	5,150	AF234648	This paper
γ2656	G	Cheyenne	6,463	AF234649	This paper
10d11	C	Cheyenne	1,213	AF234650	This paper
10h6	C[c]	Cheyenne	804	AF234651	This paper
924dc3	C[c]	Cheyenne	784	AF234645	This paper
07c8	C[c]	Cheyenne	800	AF234642	This paper
09a8	C	Cheyenne	1,026	AF234643	This paper
09df2	C	Cheyenne	1,255	AF234644	This paper

[a] G=genomic, C=cDNA, P=PCR
[b] GENBANK accession numbers
[c] Partial coding sequences
[d] Not submitted to GenBank
[e] Tetraploid (durum) wheat
[f] Spelt wheat

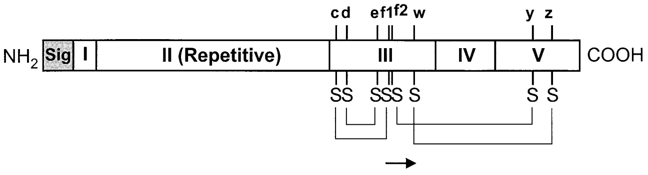


Fig. 1 Model of a γ-gliadin polypeptide. The pattern of disulfide crosslinks and cysteine residue designations are taken from Müller and Wieser (1997). The *arrow* indicates the oligonucleotide sequence used to screen cv Cheyenne genomic libraries for γ-gliadin genes. *Roman numerals* indicate major domains

C_{f2} and C_w (arrow beneath polypeptide structure in Fig. 1). Fifty 150-mm plates each containing about 15,000 plaques were hybridized with the probe. Duplicate filter lifts were carried out for each plate to distinguish false from positive signals. A single plaque was identified as reproducibly hybridizing to the γ-gliadin probe and this λ clone was named λγ13. A 1.7-kb *Hin*dIII fragment containing the hybridizing genomic fragment was subcloned into M13 and sequenced using deletion subclones (Dale et al. 1985). The λγ13 *Hin*dIII fragment was nick-translated and used as a probe to isolate additional genomic clones from the same library and other libraries described by Anderson et al. (1997).

An expressed sequence tag (EST) project identified γ-gliadin ESTs from a cv Cheyenne endosperm cDNA library. The standard M13 forward and reverse primers initiated the sequencing, followed by oligonucleotide primer walking down the clone inserts.

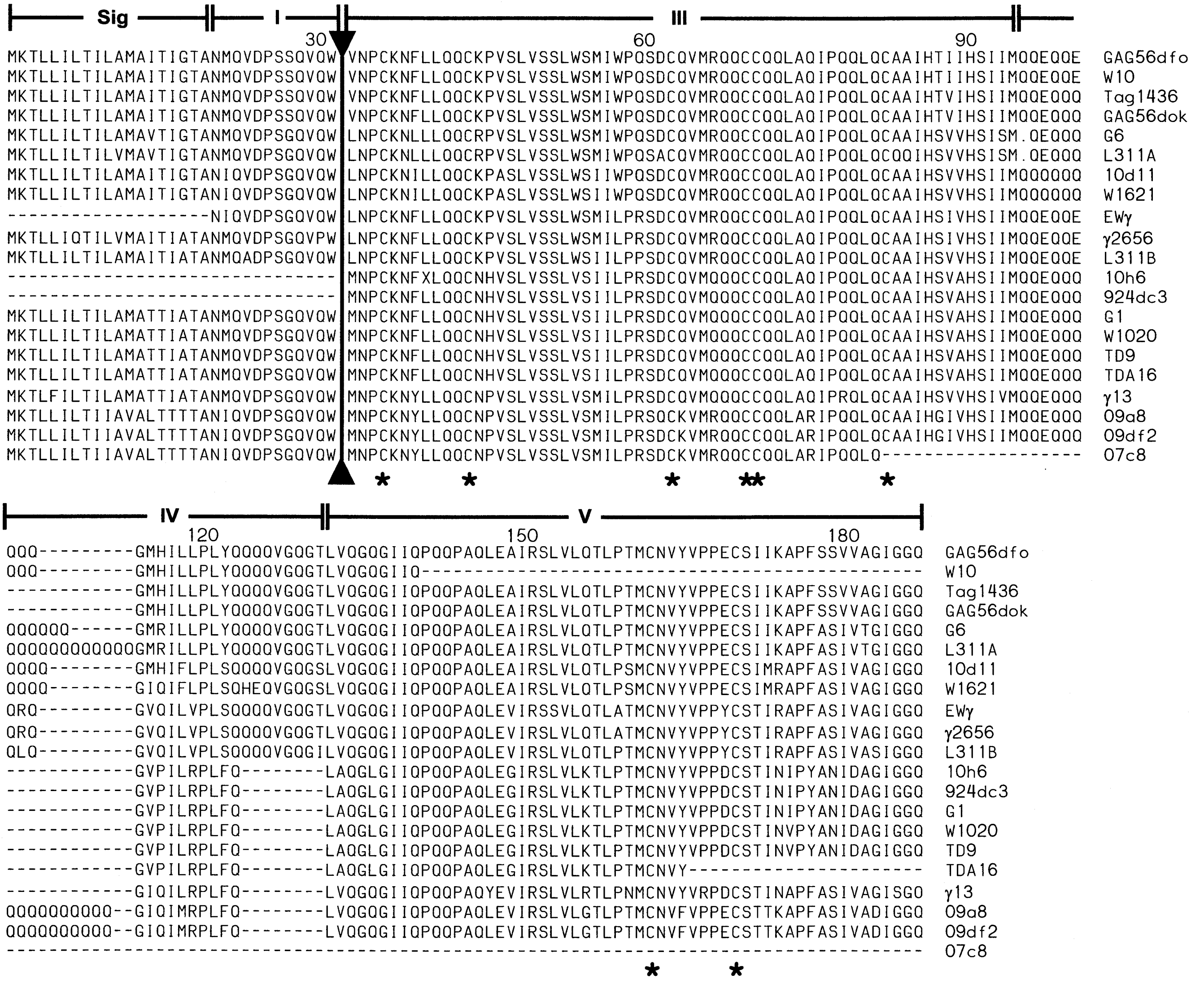
DNA sequence editing, analysis, and contig assembly were accomplished using the Editseq, Mapseq, Megalign, and SeqmanII modules of the Lasergene software (DNAstar, Inc.).

Results and discussion

A model of a γ-gliadin polypeptide is shown in Fig. 1. The 11 previously reported γ-gliadin sequences are listed in Table 1. An examination of these 11 sequences shows only six significantly different DNA sequence patterns: GAG56dfo is identical to W10, GAG56dok is only 1-bp different from Tag1436, W1020 is 3-bp different from TD9, TD9 is 1-bp different from TDA16. The wheat cDNA clones of Bartels and Thompson (1983), Okita (1984), and Okita et al. (1985) are LMW-glutenins (Cassidy et al. 1998) and not γ-gliadins as originally reported. No γ-gliadin sequences have yet been reported from the bread wheat cv Cheyenne, a high-quality bread wheat which has been a continuing focus of our effort to understand the prolamin gene families of a single model wheat cultivar.

Isolation of cv Cheyenne γ-gliadin genomic and cDNA clones

An initial genomic clone was selected from a λ phage library of wheat cv Cheyenne DNA. This clone, λγ13, was plaque-purified and *Hin*dIII fragments subcloned. A 1.7-kb *Hin*dIII fragment was sequenced and analysis found that it contained the coding and near-flanking regions of a γ-gliadin gene. This same *Hin*dIII fragment was used to re-screen wheat cv Cheyenne genomic libraries and yielded a large number of positive clones (approximately 40). Three of these were selected for further characterization based on different restriction fragment patterns (clones G1, G6, and γ2656; Table 1), and completely sequenced to include significantly more flanking DNA sequence than has previously been reported for γ-gliadin genes.



```
MKTLLILTILAMAITIGTANMQVDPSSQVQW VNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHTIIHSIIMQQEQQE GAG56dfo
MKTLLILTILAMAITIGTANMQVDPSSQVQW VNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHTIIHSIIMQQEQQE W10
MKTLLILTILAMAITIGTANMQVDPSSQVQW VNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHTVIHSIIMQQEQQQ Tag1436
MKTLLILTILAMAITIGTANMQVDPSSQVQW VNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHTVIHSIIMQQEQQQ GAG56dok
MKTLLILTILAMAVTIGTANMQVDPSGQVQW LNPCKNLLLQQCRPVSLVSSLWSMIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHSVVHSISM.QEQQQ G6
MKTLLILTILVMAVTIGTANMQVDPSGQVQW LNPCKNLLLQQCRPVSLVSSLWSMIWPQSACQVMRQQCCQQLAQIPQQLQCQQIHSVVHSISM.QEQQQ L311A
MKTLLILTILAMAITIGTANIQVDPSGQVQW LNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ 10d11
MKTLLILTILAMAITIGTANIQVDPSGQVQW LNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQCCQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ W1621
--------------------NIQVDPSGQVQW LNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQCCQQLAQIPQQLQCAAIHSIVHSIIMQQEQQE EWγ
MKTLLIQTILVMAITIATANMQVDPSGQVPW LNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQCCQQLAQIPQQLQCAAIHSIVHSIIMQQEQQE γ2656
MKTLLILTILAMAITIATANMQADPSGQVQW LNPCKNFLLQQCKPVSLVSSLWSIILPPSDCQVMRQQCCQQLAQIPQQLQCAAIHSVVHSIIMQQEQQE L311B
------------------------------- MNPCKNFXLQQCNHVSLVSSLVSIILPRSDCQVMQQQCCQQLAQIPQQLQCAAIHSVAHSIIMQQEQQQ 10h6
------------------------------- MNPCKNFLLQQCNHVSLVSSLVSIILPRSDCQVMQQQCCQQLAQIPQQLQCAAIHSVAHSIIMQQEQQQ 924dc3
MKTLLILTILAMATTIATANMQVDPSGQVQW MNPCKNFLLQQCNHVSLVSSLVSIILPRSDCQVMQQQCCQQLAQIPQQLQCAAIHSVAHSIIMQQEQQQ G1
MKTLLILTILAMATTIATANMQVDPSGQVQW MNPCKNFLLQQCNHVSLVSSLVSIILPRSDCQVMQQQCCQQLAQIPQQLQCAAIHSVAHSIIMQQEQQQ W1020
MKTLLILTILAMATTIATANMQVDPSGQVQW MNPCKNFLLQQCNHVSLVSSLVSIILPRSDCQVMQQQCCQQLAQIPQQLQCAAIHSVAHSIIMQQEQQQ TD9
MKTLLILTILAMATTIATANMQVDPSGQVQW MNPCKNFLLQQCNHVSLVSSLVSIILPRSDCQVMQQQCCQQLAQIPQQLQCAAIHSVAHSIIMQQEQQQ TDA16
MKTLFILTILAMATTIATANMQVDPSGQVQW MNPCKNYLLQQCNPVSLVSSLVSMILPRSDCQVMQQQCCQQLAQIPRQLQCAAIHSVVHSIVMQQEQQQ γ13
MKTLLILTIIAVALTTTTANIQVDPSGQVQW MNPCKNYLLQQCNPVSLVSSLVSMILPRSQCKVMRQQCCQQLARIPQQLQCAAIHGIVHSIIMQQEQQQ 09a8
MKTLLILTIIAVALTTTTANIQVDPSGQVQW MNPCKNYLLQQCNPVSLVSSLVSMILPRSDCKVMRQQCCQQLARIPQQLQCAAIHGIVHSIIMQQEQQQ 09df2
MKTLLILTIIAVALTTTTANIQVDPSGQVQW MNPCKNYLLQQCNPVSLVSSLVSMILPRSDCKVMRQQCCQQLARIPQQLQ------------------- 07c8
                                  *         *                    *      **          *

QQQ---------GMHILLPLYQQQQVGQGTLVQGQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ GAG56dfo
QQQ---------GMHILLPLYQQQQVGQGTLVQGQGIIQ----------------------------------------------- W10
------------GMHILLPLYQQQQVGQGTLVQGQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ Tag1436
------------GMHILLPLYQQQQVGQGTLVQGQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ GAG56dok
QQQQQQ------GMRILLPLYQQQQVGQGTLVQGQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFASIVTGIGGQ G6
QQQQQQQQQQQQGMRILLPLYQQQQVGQGTLVQGQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFASIVTGIGGQ L311A
QQQQ--------GMHIFLPLSQQQQVGQGSLVQGQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMRAPFASIVAGIGGQ 10d11
QQQQ--------GIQIFLPLSQHEQVGQGSLVQGQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMRAPFASIVAGIGGQ W1621
QRQ---------GVQILVPLSQQQQVGQGTLVQGQGIIQPQQPAQLEVIRSSVLQTLATMCNVYVPPYCSTIRAPFASIVAGIGGQ EWγ
QRQ---------GVQILVPLSQQQQVGQGTLVQGQGIIQPQQPAQLEVIRSLVLQTLATMCNVYVPPYCSTIRAPFASIVAGIGGQ γ2656
QLQ---------GVQILVPLSQQQQVGQGILVQGQGIIQPQQPAQLEVIRSLVLQTLPTMCNVYVPPYCSTIRAPFASIVASIGGQ L311B
------------GVPILRPLFQ--------LAQGLGIIQPQQPAQLEGIRSLVLKTLPTMCNVYVPPDCSTINIPYANIDAGIGGQ 10h6
------------GVPILRPLFQ--------LAQGLGIIQPQQPAQLEGIRSLVLKTLPTMCNVYVPPDCSTINIPYANIDAGIGGQ 924dc3
------------GVPILRPLFQ--------LAQGLGIIQPQQPAQLEGIRSLVLKTLPTMCNVYVPPDCSTINIPYANIDAGIGGQ G1
------------GVPILRPLFQ--------LAQGLGIIQPQQPAQLEGIRSLVLKTLPTMCNVYVPPDCSTINVPYANIDAGIGGQ W1020
------------GVPILRPLFQ--------LAQGLGIIQPQQPAQLEGIRSLVLKTLPTMCNVYVPPDCSTINVPYANIDAGIGGQ TD9
------------GVPILRPLFQ--------LAQGLGIIQPQQPAQLEGIRSLVLKTLPTMCNVY---------------------- TDA16
------------GIQILRPLFQ--------LVQGQGIIQPQQPAQYEVIRSLVLRTLPNMCNVYVRPDCSTINAPFASIVAGISGQ γ13
QQQQQQQQQQ--GIQIMRPLFQ--------LVQGQGIIQPQQPAQLEVIRSLVLGTLPTMCNVFVPPECSTTKAPFASIVADIGGQ 09a8
QQQQQQQQQQ--GIQIMRPLFQ--------LVQGQGIIQPQQPAQLEVIRSLVLGTLPTMCNVFVPPECSTTKAPFASIVADIGGQ 09df2
-------------------------------------------------------------------------------------- 07c8
                                                            *         *
```

Fig. 2 Alignment of all known γ-gliadin polypeptide sequences. The Clustal algorithm of Megalign (Lasergene, DNAstar, Inc.) was used to align the derived amino-acid sequences of all γ-gliadin sequences. The *vertical bar* after position 31 indicates the position of the repetitive domain (removed for this analysis). *Asterisks* indicate the positions of the eight conserved cysteine residues. *Sig*=signal peptide. Domains are as in Fig. 1

Cloning from genomic libraries is relatively time-consuming, but allows flanking DNA to be obtained. The polymerase chain reaction (PCR) is quick, but is limited to sequences similar to known sequences. Another method of gene isolation utilizes expressed sequence tag (EST) projects. In these, large numbers of random cDNA clones are sequenced. This method is relatively fast, identifies large numbers of transcribed gene sequences, and avoids relying on a single probe which can prejudice sequence selection. A potential problem with cDNA clones is that the utilized polymerase does not have proof-reading characteristics, so a low level of misincorporated bases may occur. In addition, specific cDNA libraries can be of poor quality such that artifactual duplication/deletions, frame-shifts, and stop codons are introduced. The cDNA library used in the current report evidences no such problems, and the six γ-gliadin cDNAs reported include three full-length, three truncated sequences on one end (common to cDNA clones), no stop codons, and no obvious duplication/deletions (although such occurrences within the repeat domain cannot be completely ruled out).

A preliminary screen of ESTs made from a wheat cv Cheyenne endosperm cDNA library (RNA prepared 5–30 days after flowering) indicated that approximately 1.1% of the library members hybridized strongly with a γ-gliadin probe. Sequences from a number of these clones were analyzed by BLAST and confirmed to be wheat γ-gliadin sequences. Full-length sequencing was performed on six of these clones (Table 1), three of which contained full-length coding sequences, and three of which were missing either part of the 5′ or 3′ portion of the coding regions.

Comparison of known γ-gliadin amino-acid sequences

The derived amino-acid sequences from the γ-gliadin DNA sequences listed in Table 1 are aligned in Fig. 2. The six polypeptide domains are as in Cassidy et al. (1998): a 20-residue signal peptide (Sig); a short N-terminal non-repetitive domain of the mature polypeptide (I); a highly variable repetitive domain (II; removed in Fig. 2, shown later in Figs. 3 and 4); a non-repetitive domain containing most of the cysteine residues (III); a domain rich in glutamine residues (IV); and the C-terminal non-repetitive domain containing the final two conserved cysteine residues (V). With the repetitive domain removed, the only region of length variation is domain IV, which consists of a 6–18-residue polyglutamine stretch followed nine-residues later by a nine-residue glutamine-rich region (present in about half the γ-gliadin sequences).

Polyglutamine stretches are a prominent feature in all the α-gliadins (Anderson and Greene 1997) and have previously been noted in a few members of other subfamilies of the gliadin superfamily (γ-gliadins, Rafaski 1986; LMW-glutenins, Cassidy et al. 1998). The polyglutamine region of the γ-gliadin genes is highly variable; i.e., GAG56dfo and GAG56dok differ by three CAA codons in the polyglutamine-encoding region but by only 1 bp in the remainder of their DNA sequences, and L311 A and G6 differ in the polyglutamine region by six codons and only 6 other single bp elsewhere. In many cases the DNA encoding the polyglutamine stretches can form a CAA microsatellite. Within the polyglutamine encoding regions most non-glutamine codons are 1 bp removed from glutamine codons and most likely arose from single base changes in the original glutamine codons.

The degree of sequence conservation, as seen in Fig. 2, is notable since the γ-gliadins have been considered to be the most ancient of the wheat prolamin families (Shewry and Tatham 1990). Although the α-gliadin (Anderson and Greene 1997) and LMW-glutenin (Cassidy et al. 1998) families are also internally conserved, the γ-gliadin family is most conserved (outside of the repetitive Domain II). However, ancestry can be obscured by the tendency of gene families to homogenize their sequences through mechanisms such as duplication/deletions, specific crossovers, and gene conversion.

Number and placement of cysteine residues

The wheat prolamins exist as monomers and as crosslinked polymeric proteins. Polymers range in size from a small number of crosslinked polypeptides to polymers of undetermined large size but among the largest proteins in nature (Wrigley 1996). The standard definition has been that among the wheat prolamins, the high- and low-molecular-weight glutenins form the gluten polymer while the gliadins are mainly monomeric (Shewry and Tatham 1990).

The core eight cysteines that form intramolecular disulfide linkages (Müller and Wieser 1997) are conserved in all known γ-gliadins (Fig. 2). Clone W1621 was originally reported as missing the second C (the C_d of Müller and Wieser 1997) and thus contained only seven cysteines, leaving one available for intermolecular disulfide bond formation. However, a re-sequencing on this clone found that W1621 does contain all eight conserved cysteines (Galili, personal communication; reported by Shewry and Tatham 1997). Although most γ-gliadins contain only these eight cysteine residues that form four intramolecular disulfide bonds, there are exceptions; i.e., γ13 has an additional repeat domain cysteine (see Figs. 3 and 4). In addition, W1020, TDA16, TD9 and G1 are similar to one another and have an additional cysteine in the initial part of the repeat domain (see Fig. 4), but at a different position from γ13. Thus, at least five of the 21 known γ-gliadin sequences contain an odd number of cysteines (probably more, since clones 924dc3 and 10h6 are similar to G1 but are partial sequences with the first part of the repeat domain not known).

```
CCA CAA CAA CAA
CCA TTC CGC CAG CCC CAA CAA
CCA TTC TAC CAG CAA CCA CAA CAC
ACA TTC CCC CAA CCC CAA CAA
ACA TGC CCC CAT CAA CCA CAA CAA
CAA TTT CCC CAG CCC CAG CAA CCA CAA CAA
CCA TTT CCG CAG CCC CAG CAA CCA CAA CAA
CCA TTT CCC CAG CCC CAA CAA GCC CAA CTA
CAA TTT CCC CAA CAA CCA CAA CAA
CCA TTG CCC CAG CCT CAA CAA CCC CAA CAA
CCA TTT CCC CAG TCA CAG CAA CCA CAA CAA
CCT TTT CCC CAG CCC CAA CAA CCG CAA CAA
TCA TTC CCC CAG CAA CAA CAA
CCG TTG ATT CAG
CCA TAT CTA CAA CAA CAG
```

$$\mathrm{CCA\ TT\overset{T}{C}\ CCC\ CAG\ CAA_{0-1}(CCN\ CAA_2)_{1-2}}$$

$$\mathrm{P\quad F\quad P\quad Q_{1-2}\quad (PQQ)_{1-2}}$$

Fig. 3 Alignment of repetitive motifs by DNA sequence. Repeat motif codons of the repetitive domain encoding region of clone γ13 are arrayed *vertically* to better show the relationships of the DNA structure. Proline codons are *underlined*. The single cysteine codon is *boxed*. A consensus codon repeat and the derived amino-acid residue pattern are given *below the line*

Lew et al. (1992) and Masci et al. (1995) have reported that a substantial portion of the lower-molecular-weight polypeptides in the glutenin polymer are α- and γ-type gliadin sequences. More detailed examination is needed to determine the exact nature and significance of the α- and γ-gliadins' participation in the glutenin polymer and the functional effects. Kasarda (1989) has theorized that gliadins with an odd number of cysteines would thus have one free cysteine after intramolecular bonds form. Such gliadins could participate in the gluten polymer and effectively serve as polymer terminators. We now note that approximately a quarter of the γ-gliadins can contain an uneven number of cysteine residues, and therefore the gliadins may be more of a factor in polymer formation than previously appreciated.

Repeat structure

The repetitive domain of the gliadins is composed of short peptide motifs. The exact composition of the

```
  γ13               L311B             EWγ, γ2656        09df2, 07c8,
                                                        09a8
  PQQQ              PQQQ              PQQQ
  PFRQPQQ           PFLQPHQ           PFPQPHQ           PQQQQ
  PFYQQPQH          PFSQQPQQ          PFSQQPQQ          PFPQPQQ
  TFPQPQQ           IFPQPQQ           TFPQPQQ           PFSQQPQQ
* TCPHQPQQ          TFPHQPQQ          TFPHQPQQ          IFPQPQQ
  OFPQPQQPQQ        QFPQPQQPQQ        QFSQPQQPQQ        TFPHQPQQ
  PFPQPQQPQQ        QFLQPRQ           QFIQPQQ           AFPQPQQ
  PFPQPQQAQL        PFPQQPQQ          PFPQQPQQ          TFPHQPQQ
  PFPQQPQQ          PYPQQPQQ          TYPQRPQQ          QFPQPQQPQQ
  PLPQPQQPQQ        PFPQTQQPQQ        PFPQTQQPQQ        PFPQQPQQ
  PFPQSQQPQQ        PFPQSKQPQQ        PFPQSQQPQQ        QFPQPQQPQQ
  PFPQPQQPQQ        PFPQPQQPQQ        PSPQPQQ           PFPQQPQQ
  SFPQQQQ           SFPQQQP           QFPQPQQPQQ        QFPQPQQPQQ
  PFIQ              SLIQ              SFPQQQP           PFPQLQQPQQ
  PYLQQQ            QSLQQQ            SLIQ              PLPQPQQPQQ
                                      QSLQQQ            PFPQQQQ
                                                        PLIQ
                                                        PYLQQQ

  Tag1436, W10,     G1, (924dc3)      W1020, (TDA16),   W1621, 10d11
  GAG56dfo/ok                         (TD9)
                    PQQQ                                LQQQ
  PQQQ              PFPQPQQ           PQQQ              LVPQLQQ
  PQPQPHQ         * PFCQQPQQ          PFPQPQQ           PLSQQPQQ
  PFSQQPQQ          TIPQPHQ         * PFCQ/EQPQR        TFPQPQQ
  TFPQPQQ           TFHHQPQQ          TIPQPHQ           TFPHQPQQ
  TFPHQPQQ          TFPQPQQ           PIHHQPQQ          QVPQPQQPQQ
  QFPQPQQPQQ        TYPHQPQQ          TFPQPQ/EQ         PFLQPQQ
  OFLQPQQ           QFPQTQQPQQ        TYPHQPQQ          PFPQQPQQ
  PFPQQPQQ          PFPQPQQ           QFPQTQQPQQ        PFPQTQQPQQ
  PYPQQPQQ          TF(PQQPQL         PFPQPQQ           PFPQQPQQ
  PFPQTQQPQQ        PFPQQPQQ          TFPQQPQL          PFPQTQQPQQ
  LFPQSQQPQQ        PFPQPQQPQQ        PFPQQPQQ          PFPQQPQQ
  OFSQPQQ           PFPQSQQPQQ        PFPQPQQPQQ        PFPQTQQPQQ
  OFPQPQQPQQ        PFPQPQQ           PFPQSQQPQQ        PFPQLQQPQQ
  SFPQQQP           QFPQPQQPQQ        PFPQPQQ           PFPQPQQ
  PFIQ              SFPQQQQ           QFPQPQQPQQ        QLPQPQQPQQ
  PSLQQQ            PAIQ              SFPQQQQ           SFPQQQR
                    SFLQQQ)           PAIQ              PFIQ
                                      SFLQQQ            PSLQQQ

  L311A             G6                10h6

  PQQQ              PQQQ              . . .
  PVLLPQQ           PVLLPQQ           PFPQPQQ
  PFSQQPQQ          PFSQQPQQ          TFPQQPQL
  TFPQPQQ           TFPQPQQ           PFPQQ
  TFPHQPQQ          TFPHQPQQ          PFPQPQQ
  QFPQPQQPQQ        QFP.PQQPQQ        QFPQPQQPQQ
  QFLQPQQ           QFLQPQQ           SFPQQQQ
  PFPQQPQQ          PFPQQPQQ          PAIQ
  PYPQQPQQ          PYPHQPQQ          SFLQQQ
  PFPQTQQPQQ        PFPQTQQPQQ
  LFPQSQQPQQ        LFPQSQQPQQ
  PYPQQPQQ          PYPQQPQQ
  PFPQTQQPQQ        PFPQTQQPQQ
  QFPQSQQPG.        QFPQSQQPQQ
  PFPQPQQPQQ        PFPQPQQPQQ
  SFPQQQP           SFPQQQP
  SFIQ              SFTQ
  SLQQQ             PSLQQQ
```

Fig. 4 Repeat structure of all reported γ-gliadins. The amino-acid repeat domain of all reported γ-gliadin sequences are arrayed *vertically* to compare this domain among protein family members. The identical, or nearly identical, domains are *consolidated* for brevity. The name of clone 09a8 is *underlined* along with a portion of the repeat domain to indicate that the 09a8 repeat domain is identical to 09df2 and 07c8 except that 09a8 is missing the *underlined portion* of the domain. 924dc3 is a partial sequence but has an identical repeat domain over the available sequence (*shown within parentheses*). TDA16 and TD9 (*listed in parentheses*) have identical repeat domains and are different from W1020 only at two glutamate for glutamine substitutions (W1020 residue is the first in the Q/E pairs). *Periods* indicate stop codons. *Asterisks* indicate repeat units containing cysteine residues

motifs is at least partially a subjective evaluation by individual researchers, and various consensus motifs for the γ-gliadins have been proposed: QPQQPFP (Scheets et al. 1985; D'Ovidio et al. 1995), and PQQPF plus PQQPQQ(Q)PFPQ (Rafalski 1986). D'Ovidio also considered the DNA structure and saw the DNA motif as CAA CCC CAA_2 CCA TTT CCC. Our own analyses have concentrated on the codon structure since this is the primary level of sequence change and interaction among the DNA repeat motifs (Anderson and Greene 1997; Cassidy et al. 1998). As an example, Fig. 3 shows a vertical array of the repeat structure of the γ-gliadin clone γ13. The alignment is structured to suggest ancestral codons such that a minimal number of base changes are derived; i.e., the 4th and 5th repeats begin with ACA (a threonine codon) but most likely originated as a single C to A transversion that then duplicated into a second, adjacent repeat. A proposed consensus codon and the derived amino-acid repeat motif is given at the bottom of the array.

Figure 4 shows the vertically arrayed peptide repeat pattern for all reported γ-gliadins. The initial and final repeat motifs do not fit the consensus motif but are included since they appear to be related to the consensus. The more-internal repeat motifs in the domain are undergoing relatively rapid changes mainly by single base alterations and repeat duplication/deletions. An example of a large deletion is the missing eight repeats of 09a8 as compared to 09df2 and 07c8. A shorter duplication is suggested by the repeats #9 to #11 that are duplicated in the sequence of G6.

Both G6 and L311A are similar and are pseudogenes as defined by the presence of stop codons within the repeat domain of the consensus open reading frame. With only 2 pseudogenes out of 21 sequences (13 of which are genomic sequences), the γ-gliadins are similar to the LMW-glutenin genes for which only a few pseudogenes have been reported (Cassidy et al. 1998; Lee et al. 1999) and are very different from the α-gliadins where about 50% of the genes are pseudogenes, or some zein families which can be three-quarters pseudogenes (Liu and Rubenstein 1992; Llaca and Messing 1998).

A comparison of the proposed consensus repeat motifs of all four major gliadin types is shown in Fig. 5. The LMW-glutenin consensus motif is the least similar to the other three consensus motifs. The γ-gliadin and ω-gliadin motifs are most similar, and this accounts for the cross-hybridization of a γ-gliadin probe with ω-gliadin sequences (Hsia and Anderson 2001), while the α- and γ-gliadins, and LMW-glutenins do not cross-hybridize to the same degree (Anderson et al. 1997). Such differences/similarities in repeat structure should not be used as a basis for gene-family relatedness since these repeat domains are evolving relatively rapidly, and most likely independently, within each gliadin family.

Analysis of the DNA sequence flanking coding regions

The DNA flanking the coding region of genes contains promoter and other untranslated sequences. The new γ-gliadin sequences contain standard control element motifs (data not shown) such as a TATA box and multiple polyadenylation signals that have been reported for other gliadin genes. In the current report, one objective was to obtain more distal 5´ and 3´ non-coding sequenc-

Fig. 5 Repeat domain motifs for the major classes of the gliadin superfamily. A comparison of the repeat domains within members of the different major classes of gliadins suggested different, but related, consensus repeat motifs for each class. The LMW-glutenin pattern is taken from Cassidy et al. (1998). The α-gliadin pattern is taken from Anderson and Greene (1997). The γ-gliadin pattern is from the present report, and ω-gliadin pattern is from the accompanying paper (Hsia and Anderson 2001)

Gliadin Class	Codons	Amino Acids
LMW-glutenin	$\mathrm{CCA}_{1\text{-}2}$ TTT ${}^{T}_{C}\mathrm{C}^{A}_{G}$ CA^{A}_{G} $\mathrm{CAA}_{1\text{-}5}$	$\mathrm{P}_{1\text{-}2}\mathrm{F}\ \substack{\mathrm{P}\\ \mathrm{S}}\ \mathrm{Q}_{2\text{-}6}$
α-gliadin	CCA $\mathrm{T}\substack{T\\A}\mathrm{T}$ CC^{A}_{G} $\mathrm{CA}^{G}_{A\ 0\text{-}1}$ CC^{G}_{A} $\mathrm{CAA}_{1\text{-}2}$	$\mathrm{P}\ \substack{\mathrm{F}\\ \mathrm{Y}}\ \mathrm{P}\ \mathrm{Q}_{0\text{-}1}\ \mathrm{P}\ \mathrm{Q}_{1\text{-}2}$
γ-gliadin	CCA TT^{T}_{C} CCC CAG $\mathrm{CAA}_{0\text{-}1}$ $(\mathrm{CCN}\ \mathrm{CAA}_{2})_{1\text{-}2}$	$\mathrm{P\ F\ P\ Q}_{1\text{-}2}(\mathrm{PQQ})_{1\text{-}2}$
ω-gliadin	CCA TT^{T}_{C} CCC CAG $\mathrm{CAA}_{0\text{-}1}$ CC^{C}_{T} CAA_{2}	$\mathrm{P\ F\ P\ Q}_{1\text{-}2}\mathrm{PQQ}_{1\text{-}2}$

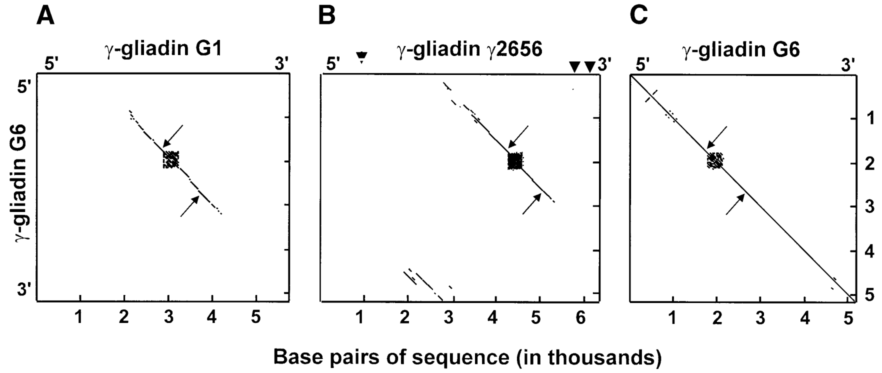


Fig. 6A–C Homology plots of γ-gliadin sequences. Pairs of γ-gliadin DNA sequences show homologies between genes using a criterion of an 80% match over a 20-bp window. **A** G6 vs G1. **B** G6 vs γ2656. **C** G6 vs G6. The *box-like* densely plotted region on each diagonal indicates the multiple homologies among the repetitive domain motifs. *Arrows above the diagonals* indicate positions of start codons and *arrows below the diagonals* indicated stop codons. *Arrowheads* along the γ2656 sequence indicate the position of short regions homologous to cereal repetitive DNAs. The 5′ and 3′ ends of the sequences are indicated on the left and top axis

es than previously reported. Three clones, G1, G6 and γ2656, were sequenced to obtain 4–5 kb of flanking sequence. Figure 6 shows a pairwise comparison of these more-extensive sequences. G1 and G6 flanking DNA diverge beyond about –700 upstream with respect to the initiation codon and +600 downstream with respect to the stop codon. Clones G6 and γ2656 diverge similarly in their 5′ regions, but γ2656 diverges at approximately 300 bp downstream from the stop codon from all known γ-gliadin sequences. Among the previously characterized γ-gliadin genes, the most-flanking DNA (600–1,000 bp for both 5′ and 3′ regions) has been reported for LP311 A and LP311B, and both genes have similar flanking sequences to G6. Not enough sequence data is currently available for definitive conclusions, but the patterns of divergence are similar to that reported for the α-gliadin gene family (Anderson et al. 1997) wherein the the α-gliadin flanking DNA sequences diverged at approximately –600 bp at the start codon about 500 bp 3′ to the polyadenylation sites. We had previously speculated that the conserved regions are the delimiting DNA sequence elements needed for fully functional gliadin genes, on the assumption that genes missing elements necessary for gene control would be inactive and not maintained through selection.

The G6 vs γ2656 comparison (Fig. 6) shows a second region of similarity (seen as the lower short diagonal) implying that the 3′ portion of the G6 sequence is similar, but not identical, to part of the 5′ flank of γ2656. These similar sequences in the flanking DNAs could be repetitive elements occurring by chance near these two γ-gliadin genes or it could indicate that G6 is immediately upstream of another γ-gliadin gene. Among the known γ-gliadins, G6 is most similar to L311A, the 5′ member of a pair of γ-gliadin genes found on a single λ clone (Rafalski 1986). L311A is followed 4-kb downstream by a second γ-gliadin gene, L311B, that is most similar to γ2656. Not enough genetic or physical mapping data is available to determine if the Cheyenne γ-gliadin genes G6 and γ2656 are allelic to the linked γ-gliadin genes of Rafalski (1986) or if close linkages are a more general characteristic of the γ-gliadin gene family.

There were few significant matches between the γ-gliadins flanking DNA sequences and sequences in the public databases. The exception was for γ2656 (arrowheads along the γ2656 sequence in Fig. 6) where both the 5′ and 3′ flanks have short homologies to cereal repetitive DNAs. In the 5′ flank of γ2656 is a 50-bp sequence with low homologies (e^{-7} to e^{-10}) to sequences found in sequenced rice and sorghum BAC (bacterial artificial chromosome) clones. In the 3′ flank, and just following the polyadenylation site, is a 33-bp fragment with low

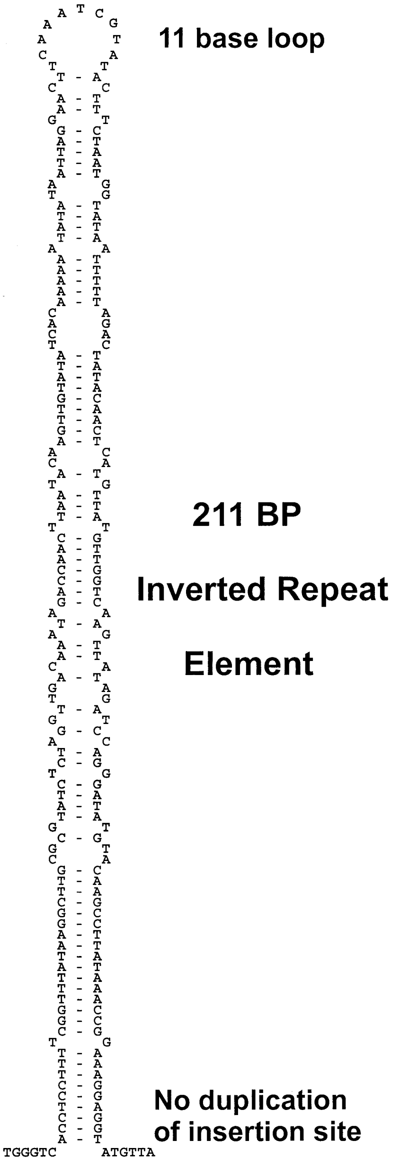


Fig. 7 Proposed structure of a MITE-like element found in the G6 5′ flanking DNA. The potential snap-back secondary structure of a 211-bp DNA sequence is given. Complementarity is indicated by *dashes*. The short, nonduplicated, sequence adjacent to both ends of the suggested MITE-like element is shown at the base of the figure

homology to a rye hypervariable DNA (Genbank accession AF153326) and part of a wheat retrotransposon WIS-2–1A (Lucas et al. 1992). Further downstream is a longer segment (202 bp) with significant homologies to a *Hordeum chilense* repeated DNA (Ferrer et al. 1995; $8e^{-34}$) and a repeated sequence reported from *Psathyrostachys juncea* (McIntyre et al. 1988; e^{-44}). Outside of these exceptions, most available γ-gliadin flanking DNA has no significant homology to any known sequences, consistent with observations made with other wheat prolamin gene families (Anderson, data not shown). These sequences may be non-informative DNA specific to wheat, but further genomic sequencing within the grasses will clarify such issues of genome sequence evolution.

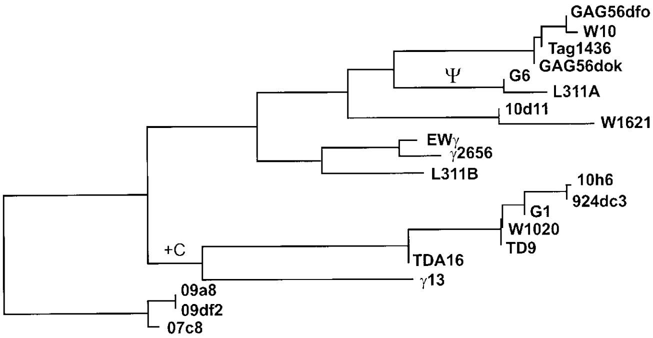


Fig. 8 Phylogeny tree of γ-gliadins. The alignment from Fig. 2 was used to construct a phylogenetic tree of all reported γ-gliadin sequences. The *+C* indicates a branch all of whose known members contain an extra cysteine residue. The Ψ *symbol* indicates a branch whose members are pseudogenes

A dotplot analysis of G6 with itself (right frame in Fig. 6) shows a short perpendicular diagonal through the 5′ flank sequence diagonal. Such a “cross” is indicative of an inverted repeat forming a MITE (miniature inverted transposable element), although in this case the element is missing the duplicated insertion site typical of transposable element insertion events. The proposed self-annealing of this DNA is shown in Fig. 7. This is the only such MITE-like element we have found near a gliadin gene, but we have identified several in the 5′ regions of the HMW-glutenin genes (data not shown).

Gliadin flanking sequences have yet to be associated with regions of nested transposons as found in maize (SanMiguel et al. 1996). Either the wheat genome is not constructed the same as maize, or such repeat nests have not yet been reached in available wheat flanking sequences. Further study of gliadin loci structure using bacterial artificial chromosome libraries should provide clearer information on the extended physical structure of these loci.

Phylogenetic tree of known γ-gliadin sequences

Sabelli and Shewry (1991) estimated 15–40 different γ-gliadin genes in the hexaploid cv Chinese Spring. We believe the number in cv Cheyenne is likely to be close to the high end of the Sabelli and Shewry estimate since 10 of the 12 Cheyenne sequences initially determined were unique. Two cDNA sequences were identical to cDNA 10d11, but identity does not establish origin from the same gene because of the possibility of recent duplication/conversions in such complex gene families.

Figure 8 shows a phylogenetic tree derived from the alignment in Fig. 2, and indicates clustering of the known γ-gliadin sequences. Most of the new sequences reported here fall within existing clusters, except for three of the cDNA sequences (09a8, 09df2 and 07c8) which form a previously unreported subgroup of the γ-gliadin gene family.

A general problem with studies of large gene families is that the identification of additional family members can depend strongly on the initial probe. For example, if the first isolated member of a family is used to screen a genomic library, then there will be the tendency to isolate new genes most closely related to the probe sequence. Similarly, use of PCR to avoid library screening depends on knowledge of previously studied genes. In the case of the γ-gliadin sequences, all of the previously reported sequences, as well as the four new genomic clones in the present report, have a direct or indirect descent from the first γ-gliadin partial cDNA clone, W10 (Scheets et al. 1985). A random cDNA screen (such as the one revealing the 09a8, 09df2 and 07c8 cDNA sequences) has an advantage in avoiding these biases. Our previous understanding of the structure of the prolamin gene families will change with large-scale EST programs identifying new and previously unknown gene family structures and relationships.

Acknowledgments The authors thank Renato D'Ovidio, Don Kasarda, and Pat Okubara for reading the manuscript and making suggestions and corrections. All experiments comply with U.S. laws.

References

Anderson OD, Greene FC (1997) The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59–65

Anderson OD, Halford NG, Forde J, Yip R, Shewry PR, Greene FC (1988) Structure and analysis of the high-molecular-weight glutenin genes from *Triticum aestivum* cv Cheyenne. In: Miller TE, Koebner RMD (eds) Proc 7th Int Wheat Genet Symp. Bath Press, pp 699–704

Anderson OD, Litts JC, Greene FC (1997) The α-gliadin gene family. I. Characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of the gene family. Theor Appl Genet 95:50–58

Bartels D, Thompson RD (1983) The characterization of cDNA clones coding for wheat storage proteins. Nucleic Acids Res 11:2961–2977

Bartels D, Altosaar I, Harberd NP, Barker RF, Thompson RD (1986) Molecular analysis of γ-gliadin gene families at the complex *Gli-1* locus of bread wheat (*T. aestivum* L.). Theor Appl Genet 72:845–853

Büren M von, Lüthy J, Hübner P (2000) A spelt-specific γ-gliadin: discovery and detection. Theor Appl Genet 100:271–279

Cassidy BG, Dvorak J, Anderson OD (1998) The wheat low-molecular-weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor Appl Genet 96:743–750

Dale RMK, McClure BA, Houchins JP (1985) A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 s rDNA. Plasmid 13:31–40

D'Ovidio R, Simeone M, Masci S, Porceddu E, Kasarda DD (1995) Nucleotide sequence of a γ-type glutenin gene from a durum wheat: correlation with a (-type subunit from the same biotype. Cereal Chem 72:443–449

D'Ovidio R, Tanzarella OA, Porceddu E (1991) Cloning and sequencing of a PCR-amplified gamma-gliadin gene from durum wheat [*Triticum turgidum* (L.) Thell]. Plant Sci 75:229–236

Ferrer E, Loarce Y, Hueros G (1995) Molecular characterization and chromosome location of repeated DNA sequences in *Hordeum* species and in the amphiploid tritordeum (×*Tritordeum* Ascherson et Graebner). Genome 38:850–857

Hsia CC, Anderson OD (2001) Isolation and characterization of wheat ω-gliadin genes. Theor Appl Genet (in press)

Kasarda DD (1989) Glutenin structure in relation to wheat quality. In: Pomeranz Y (ed) Wheat is unique. American Association of Cereal Chemists, pp 277–302

Kreis M, Shewry, PR, Forde BG, Forde J, Miflin BJ (1985) Structure and evolution of seed storage proteins and their genes with particular reference to those of wheat, barley, and rye. Oxford Surveys Plant Mol Cell Biol 2:253–317

Lee Y-K, Ciaffi M, Appels R, Morell MK (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. II. The genes from A-genome species. Theor Appl Genet 98:126–134

Lew E J-L, Kuzmicky DD, Kasarda DD (1992) Characterization of low-molecular-weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino-acid sequencing. Cereal Chem 69:508–515

Liu CN, Rubenstein I (1992) Molecular characterization of two types of 22-kilodaltons alpha-zein genes in a gene cluster in maize. Mol Gen Genet 234:244–253

Llaca V, Messing J (1998) Amplicons of maize zein genes are conserved within genic but expanded and constricted in intergenic regions. Plant J 15:211–220

Lucas H, Moore G, Murphy G, Flavell RB (1992) Inverted repeats in the long-terminal repeats of the wheat retrotransposon Wis-2–1A. Mol Biol Evol 9:716–728

Maruyama N, Ichise K, Katsube T, Kishimoto T, Kawase S, Matsumura Y, Takeuchi Y, Sawada T, Utsumi S (1998) Identification of major wheat allergens by means of the *Escherichia coli* expression system. Eur J Biochem 255:739–74

Masci S, Lew EJ-L, Lafiandra D, Porceddu E, Kasarda DD (1995) Characterization of low-molecular-weight glutenin subunits in durum wheat by reversed-phase high-performance liquid chromatography and N-terminal sequencing. Cereal Chem 72:100–104

McIntyre CL, Clarke BC, Appels R (1988) Amplification and dispersion of repeated DNA sequences in the Triticeae. Plant Syst Evol 160:39–59

Müller S, Wieser H (1997) The location of disulphide bonds in monomeric (-type gliadins. J Cereal Sci 26:169–176

Okita TW (1984) Identification and DNA sequence analysis of a γ-gliadin cDNA plasmid from winter wheat. Plant Mol Biol 3:325–332

Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of the α/β-type and γ-type gliadin DNA sequences. J Biol Chem 260:8203–8213

Payne PI (1987) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. Annu Rev Genet 38:141–153

Rafalski JA (1986) Structure of wheat gamm-gliadin genes. Gene 43:221–229

Sabelli PA, Shewry PR (1991) Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheats by restriction fragment analysis. Theor Appl Genet 83:209–216

SanMiguel P, Tikhonov A, Jin Y-K, Motchaoulskala N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avrarnova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:75–768

Scheets K, Hedgcoth C (1988) Nucleotide sequence of a γ gliadin gene: comparisons with other γ gliadin sequences show the structure of γ gliadins genes and the general primary structure of γ gliadins. Plant Sci 57:141–150

Scheets K, Rafalski JA, Hedgcoth C, Söll DG (1985) Heptapeptide repeat structure of a wheat γ-gliadin. Plant Sci 37:221–225

Shewry PR (1995) Plant storage proteins. Biol Rev 70:375–426

Shewry PR, Tatham AS (1990) The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J 267:1–12

Shewry PR, Tatham AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25:207–227

Singh NK, Shepherd KW (1988) Linkage mapping of genes controlling endosperm storage proteins in wheat. 1. Genes on the short arms of group-1 chromosomes. Theor Appl Genet 75:628–641

Sugiyama T, Rafalski A, Söll D (1986) The nucleotide sequence of a wheat γ-gliadin genomic clone. Plant Sci 44:205–209

Wrigley CW (1996) Giant proteins with flour power. Nature 381:738–739