Molecular Characterization and Evolutionary Analysis of α-gliadin Genes from *Eremopyrum bonaepartis* (Triticeae)

Guangrong Li, Tao Zhang, Yirong Ban & Zujun Yang (Corresponding author)

School of Life Science and Technology, University of Electronic Science and Technology of China

Chengdu 610054, Sichuan, China

Tel: 86-28-8320-6556   E-mail: yangzujun@uestc.edu.cn

Abstract

Total 16 α-gliadin gene sequences ranged from 1186 to 1316bp were isolated from *Eremopyrum bonaepartis* by PCR based strategy. Analysis of deduced amino acid sequences indicated that 8 of 16 sequences displayed the typical structure of α-gliadin genes with six cysteine residues, while the other 8 sequences contained in-frame stop codon, and were therefore pseudogenes. Phylogenetic analysis based on the variation of α-gliadin gene sequences indicated that the F genome of *E. bonaepartis* was apparently differed from the A, B and D genomes of wheat. Four peptides, Glia-α, Glia-α2, Glia-α9 and Glia-α20, which have been identified as T cell stimulatory epitopes in celiac disease (CD) patients through binding to HLA-DQ2/8, were searched to the *E. bonaepartis* α-gliadin gene sequences. We firstly found that only Glia-α existed in 7 of 8 full sequences, the fact suggesting that the occurrence of the stimulatory epitopes in α-gliadin genes was newly evolved in wheat.

Keywords: *Eremopyrum bonaepartis*, α-gliadin, Celiac disease epitopes, Triticeae

1. Introduction

The total protein contents of cereal seeds vary from about 10–15% of the grain dry weight, with about half of the total being storage proteins. The majority of seed storage proteins are prolamins. All of the prolamins of the Triticeae (wheat, barley and rye) were assigned to three broad groups: sulphur-rich (S-rich), sulphur-poor (S-poor) and high molecular weight (HMW) prolamins (Shewry and Tatham, 1990). The S-rich prolamins included subgroups of γ-gliadins, α/β-gliadins and B- and C-type LMW subunits of glutenin. Examination of the number of gene loci (*Gli*-2) for α/β-gliadins type subfamilies showed that the copy number ranges from 25 to 150 copies in wheat (*Triticum aestivum* L), and such variation were presumably caused by genes differential amplification during genome evolution (Anderson and Greene, 1997; Gu et al., 2004; Gao et al., 2007).

The Triticeae tribe is a taxon within the grass family Poaceae that includes around 350 species in approximately 30 genera (Löve 1984). The genus *Eremopyrum* (Ledeb.) Jaub. & Spach was widely distributed in dry areas from Morocco to western China (Frederiksen, 1991). As far as the *Eremopyrum* species was concerned, studies are focused on their morphology, conservation and genomic relationship to related diploid and polyploidy triticeae species (Frederiksen and Bother, 1994; Kellogg et al., 1996; Liu and Ding, 1996; Mason-Gamer, 2004). In this report, we describe the characterization of the α-gliadin genes from *Eremopyrum bonaepartis*, an important diploid pasture species (2x=14), with respect to explore the evolution of α-gliadin genes in triticeae.

2. Materials and Methods

2.1 Plant materials

Eremopyrum bonaepartis, accession PI401345 was kindly provided by Dr. Harold Bockelman, National Plant Germplasm System, USDA-ARS, Aberdeen, Idaho, U. S. A.

2.2 DNA preparation and PCR amplification

Genomic DNA of *E. bonaepartis* was isolated from 1-2g leaves of single plant about 4-6 weeks old, as described by Li et al (2008). Based on the published α-gliadin gene sequence, primers P1 (5'-CCACAAATCCAACATG(A/T)AGACCTT(C/T)(A/C)T   -3'), and P2 (5'-TTCCCATG(C/T)TTGAACTA(G/C)TATAGGT(C/A)G -3') were designed for amplifying the upstream sequence, coding sequences and a downstream sequences of α-gliadin genes.

PCR was carried out using iCycler thermal cycler (Bio-Rad, Hercules, CA). Each PCR reaction (100μl) contained 300ng template, 0.2mM of each dNTPs, 1μM of each of two primers, 10μl 10×PCR buffer, and 5U high fidelity Ex Taq DNA polymerase (Takara, Japan). The PCR programmed at 2min at 95°C, followed by 35 cycles of 94°C for 1min, 60°C for 1min and 72°C for 2min 30sec. After the amplification, the final extension was kept for 10min at 72°C. The PCR product was analyzed on 1% agarose gels.

2.3 Cloning, sequencing and comparative analyses of α-gliadin genes

The target PCR products were purified from agarose gel using QIAquick Gel Extraction Kit (QIAGEN). The fragment was ligated into a pGEMT plasmid vector (Promega, Madison, Wis.), and used to transform competent cell of *Escherichi coli* DH5α strain. The 15 positive clones were sequenced on an automatic DNA sequencer (TaKaRa Biotech). The nucleotide sequences were assembled, and the sequence alignment among different α-gliadin alleles from NCBI Genbank were carried out using BIOEDIT software, and the phylogenetic tree was constructed by ClustalW program.

3. Results

3.1 Nucleotide sequences analysis

By using the *E. bonaepartis* genomic DNA as templates, the amplification of primers, P1/P2, gave rise to a 1200bp band. The PCR products of *E. bonaepartis* were cloned and sequenced. Total 16 sequences of 1186-1316bp were obtained and deposited in the NCBI database with accession number of HM452979 to HM452994.

Nucleotide sequences comparison indicated that all 16 sequences showed a high degree of homology to the wheat α-gliadin sequences. Eight sequences (HM452979 to HM452986) contained complete ORFs of 834-969bp, while the remaining 8 sequences (HM452987 to HM452994) with structurally similar to the full-ORF genes were pseudo genes, because they contained a typical in-frame premature stop codons resulted from the transition of T by C at the first base of glutamine codon (CAA, CAG) in either repetitive central domains or C- terminal domains.

The 16 sequences contained the conserved 214-217bp 5' upstream sequences. As shown in Figure 1, the sequences HM452982, HM452986 and HM452987 showed higher nucleotide variations than other 13 sequences. The TATA box was commonly found at 92-96bp upstream of the ATG start codon. Several predicted cis-regulatory elements like CAAT and the endosperm specific GCN4 presented in the promoter regions of the sequences. The circadin factors existed in 13 sequences.

3.2 The deduced amino acid sequences analysis

The deduced amino acid sequence of the *E. bonaepartis* α-gliadin genes represented a presumptive mature protein with 277-323 residues (Figure 2). All of full-ORFs are resembled the general structure of α-gliadin protein, which consists of a short N-terminal signal peptide (S) followed by a repetitive domain (R) and a longer non-repetitive domain (NR1 and NR2), separated by two polyglutamine repeats (Q1 and Q2). The deduced amino acid of the *E. bonaepartis* α-gliadin genes contained six conserved cystenin residues in the non-repetitive domains and C-terminal domains. The two polyglutamine domains were encoded by microsatellite-like sequences, CAA and CAG. As shown in Figure 2, the repeats numbers of Q1 and Q2 in the ORF were 9-10, and 14-28, respectively.

The reported T cell stimulatory epitopes Glia-α (QGSFQPSQQ), Glia-α2 (PQPQLYPQ), Glia-α9 (PFPQPQLPY) and Glia-α20 (FRPQQPYPQ) had its own position in the α-gliadin protein, in which Glia-α was present in the second nonrepetitive (NR2) domain, whereas Glia-α2, Glia-α9 and Glia-α20 were all found in the first repetitive (R) domain (Spaenij-Dekking et al., 2005). We search the perfect matches in the obtained *E. bonaepartis* α-gliadin genes sequences to the four epitopes, and found that only the Glia-α present in 7 of 8 *E. bonaepartis* α-gliadin ORF sequences. The changes of several amino acids were at position of other 3 epitopes regions.

3.3 Phylogenetic tree and evolutionary analyses

Sequence comparisons were performed among the α-gliadin genes from different genomes to understand the relatedness and the divergent time by the construction of phylogenetic trees. In addition to the nucleotide sequence of α-gliadins alleles from A, B and D genome from wheat ancestries (van Herpen et al., 2006), the HMW-GS, LMW-GS, γ-gliadins gene sequences were also included. As shown in Fig. 2, the γ-gliadins were closely related to α-gliadins. Among the α-gliadin gene sequences, the phylogenetic trees of *E. bonaepartis* α-gliadin gene sequences were apparently separated as three distinguished groups. Based on the calculation of the evolutionary rates **(**Allaby et al., 1999), the α-gliadin sequences from *E. bonaepartis* possibly evolved at

15-20MY, while the generation of the A, B, D genome α-gliadin sequences at 8-10MY in their evolutionary history (Fig. 2).

4. Discussion

Since the complete amino acid sequence from α-gliadin in wheat was reported by Kasarda et al. (1984), the general conserved amino acid MKTFLIL and FGIFGTN of encoding sequences were found, and the several primers were used to amplify the α-gliadin genes from *Triticum* species (Kasarda and D'Ovidio, 1999; van Herpen et al., 2006). Seed storage protein gene promoter motifs are conserved in diverse species, and different plant families are characterized by a distinct combination of conserved motifs (Fauteux and Strömvik, 2009). In the present study, we designed a pair of primers, P1/P2, which covered a 214-217bp of the promoter and 3' untranslated regions of α-gliadin gene sequences. It successfully amplified the α-gliadin genes from *Eremopyrum* species. It also proved to successfully amplify the α-gliadin genes from *Dasypyrum*, *Pseudoroegneria*, *Lophopyrum*, and *Thinopyrum* species (Li et al. 2009 and unpublished data). The PCR based strategy effectively enables to characterize new types of α-gliadin gene sequences in Triticeae species. Using the PCR amplification, van Herpen et al. (2006) reported 230 partial α-gliadin gene sequences from diploid ancestral A, B, D genome of wheat, and found that about 87% of sequences are pesudogenes. We previously isolated 32 α-gliadin gene sequences from *Dasypyrum* species, and revealed that 12 sequences are pseudogenes (Li et al., 2009). In the present study, we showed that 8 of 16 α-gliadin gene sequences from *Eremopyrum* species were pseudogenes, suggesting that high percentage of pseudogenes existed in α-gliadin gene superfamily of Triticeae genomes.

There are large differences existed in the content of predicted T cell epitopes in full-ORF genes and pseudogenes from the wheat ancestral genome donor of diploid species of *T. monococcum*, *T. speltoides* and *Ae. tauschii* (Spaenij-Dekking et al., 2005; van Herpen et al. 2006; Vaccino et al. 2009). Our previous study revealed that the α-gliadin sequences from V genome of *Dasypyrum* present Glia-α epitope (Li et al., 2009). In the present study, we showed that *Eremopyrum* species only contained Glia-α epitope. It is also worthwhile to postulate the evolutionary history of the toxic epitopes by the investigation of accumulative α-gliadin gene sequences from closely related genomes.

The major groups of cereal prolamins in the Triticeae (wheat, barley, rye) and the Panicoideae (maize, sorghum, millets) have common evolutionary origins (Gianibelli et al., 2001). However, the tandem gene duplications events and complicating breeding efforts occurred in the multigenic families that encode storage proteins including α-gliadin genes (Xu and Messing, 2008). Therefore, further isolation of the α-gliadin gene sequences from different genus or species in Triticeae will resolve the ancestry of gene copies, and can trace the origin of chromosome lineages with respect to the gene evolution.

References

Allaby, R.G., Banerjee, M. & Brown, T.A. (1999). Evolution of the high molecular weight glutenin loci of the A, B, D and G genomes of wheat. *Genome*, 42, 296-307.

Frederiksen, S. & von Bothmer, R. (1994). Intergeneric hybridizations with *Eremopyrum* (Poaceae). *Nordic Journal of Botany*, 15, 39-47.

Anderson, O.D. & Greene, F.C. (1997). The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure and origins of pseudogenes. *Theor. Appl. Genet.*, 95, 59-65.

Frederiksen, S. (1991). Taxonomic studies in *Eremopyrum* (Poaceae). *Nordic Journal of Botany*, 11, 271-285.

Fauteux, F. & Strömvik, M.V. (2009). Seed storage protein gene promoters contain conserved DNA motifs in Brassicaceae, Fabaceae and Poaceae. *BMC Plant Biology,* 9, 126.

Gao, S., Gu, Y.Q., Wu, J., Coleman-Derr, D., Huo, N., Crossman, C., Jia, J., Zuo, Q., Ren, Z., Anderson, O.D. & Kong, X. (2007). Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. *Plant Mol. Biol.*, 65, 189-203.

Gianibelli, M.C., Larroque, O.R., Macritchie, F. & Wrigley, C.W. (2001). Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. *Cereal. Chem.*, 78, 635-646.

Gu, Y.Q., Crossman, C., Kong, X., Luo, M., You, F.M., Coleman-Derr, D., Dubcovsky, J. & Anderson, O.D. (2004). Genomic organization of the complex alpha-gliadin gene loci in wheat. *Theor. Appl. Genet.*, 109, 648-657.

Kasarda, D.D. & D'Ovidio, R. (1999). Deduced amino acid sequence of an α-gliadin gene from spelt wheat (Spelta) includes sequences active in celiac disease. *Cereal Chem.*, 76, 548-551.

Kasarda, D.D., Okita, T.W., Bernardin, J.E., Baecker, P.A., Nimmo, C.C., Lew, J.L., Dietler, M.D. & Green, F.C. (1984). Nucleic (cDNA) and amino acid sequences of α-type gliadins from wheat (*Triticum aestivum*). *Proc Natl Acad Sci USA*, 81, 4712-4716.

Kellogg, E.A., Appels, R. & Mason-Gamer, R.J. (1996). When genes tell different stories: the diploid genera of Triticeae. *Syst. Bot.*, 21, 321-347.

Li, G.R., Ren, Z.L., Liu, C., Zhou, J.P. & Yang, Z.J. (2008). Isolation and Sequence Analysis of α-gliadin Genes from *Dasypyrum breviaristatum*. *Acta Agronomica Sinica*, 34, 1097-1103.

Li, G.R., Liu, C., Zeng, Z.X., Jia, J.Q., Zhang, T., Zhou, J.P., Ren, Z.L. & Yang, Z.J. (2009). Identification of α-gliadin genes in *Dasypyrum* in relation to evolution and breeding. *Euphytica*, 165, 155-163.

Löve, A. (1984). Conspectus of the Triticeae. *Feddes Repertorium*, 95, 425-521.

Liu, J.W. & Ding, M. (1996). Morphological and cytogenetical studies on intergeneric hybrids between *Triticum aestivum* and *Eremopyrum orientale*. *Acta Genetica Sinica*, 23, 117-123.

Mason-Gamer, R.J. (2004). Reticulate evolution, introgression, and intertribal gene capture in an allohexaploid grass. *Syst. Biol.*, 53, 25-37.

Shewry, P.R. & Tatham, A.S. (1990). The prolamin storage proteins of cereal seeds: structure and evolution. *Biochem. J.*, 267, 1-12.

Spaenij-Dekking, L., Kooy-Winkelaar, Y., van Veelen, P., Wouter Drijfhout, J., Jonker, H., van Soest, L., Smulders, M.J.M., Bosch, D., Gilissen, L.J.W.J. & Koning, F. (2005). Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. *Gastroenterology*, 129, 797-806.

Van Herpen, T.W., Goryunova, S.V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O., Schenk, M.F., van Veelen, P.A., Koning, F., van Soest, L.J., Vosman, B., Bosch, D., Hamer, R.J., Gilissen, L.J. & Smulders, M.J. (2006). Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. *BMC Genomics* 7, 1.

Vaccino, P., Becker, H., Brandolini, A., Salamini, F. & Kilian, B. (2009). A catalogue of Triticum monococcum genes encoding toxic and immunogenic peptides for celiac disease patients. *Mol Genet Genomics*, 281, 289-300.

Xu, J. & Messing, J. (2008). Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc. Natl. Acad. Sci. U.S.A*, 105, 14330–14335.
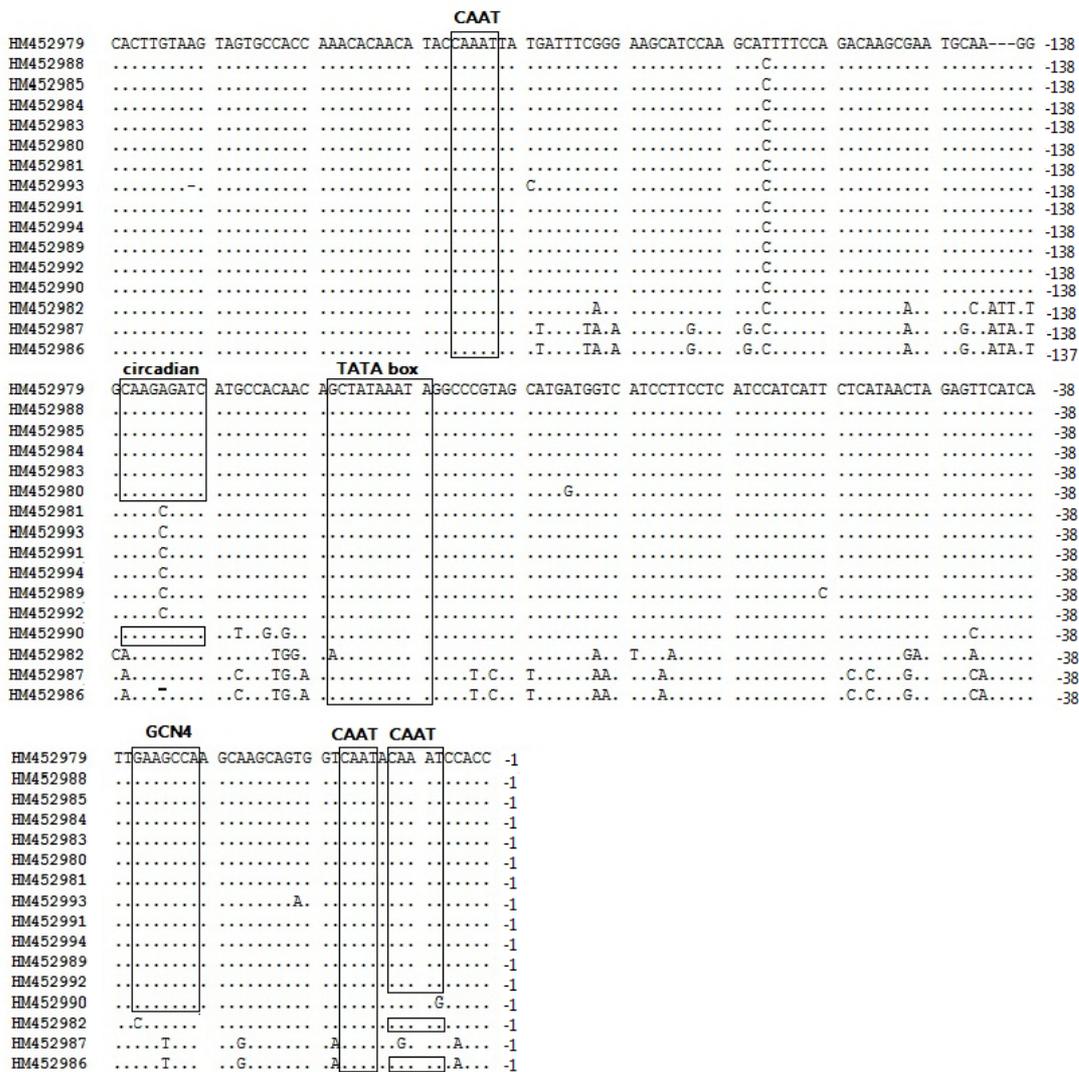
Figure 1. The promoter region of the 18 α-gliadin genes sequences. The predicted cis- regulatory elements were boxed

```
                     Signal peptide ──▶|                              Repetitive domain
        HM452979  MKTFLILALL AIVATTATTA VRVPVPQSQP QNPSQQQPQE EIPLVQQQQF PGQQQPFPPQ QPYPQPQPFP PQQPYPQPQ- -----PFSPQ  84
        HM452980  .......... .......... .......... .......... .......... .......... .......... .......... ..........  84
        HM452981  .......... .......... .......... .......... QV..M...-- -----.L... .....L.... S......... .......P--  75
        HM452982  .......S.. .......... L......L.. .......... QV..M...-- -----.L... .......... S......... .......P--  75
        HM452983  .......... ......S... .......... .......... .......... .......... .......... .......... ......S....  84
        HM452984  .......... ......T... .......... .......... .......... .......... .......... .......... ..........  84
        HM452985  .......... .......... .......... .......... .V........ .......S.. .......... .......... ..........  84
        HM452986  .......... ......T... ..G....M.. ....L....D QS..M...-. ..K....... .S......LS .........P VPPQQ.YPQP  89
                                                                    |◀──────  Q1  ──────▶|
        HM452979  QPFLQPQPFP PQLPYPQPQP FPPQQPYPQP QQQYPQPQQP ISQQ---QAQ QILQPLLQQQ LIP[C]MDVVLQ QHNIAHASSQ VLQQSSYQEL 171
        HM452980  .......... .......... .......... .......... .......... .......... .......... .......... .......... 171
        HM452981  ----...... .......... ..T....... .........E. .L........ ....QFQ... ...[R].... ..S........ .....-..G. 157
        HM452982  ----...... .......... ..T....... .........Q...E. .......... ....QFQ... ...[R].... ..S..----- .....-..G. 152
        HM452983  .......... .......... .......... .......... .......... .......... .......... .......... .......... 171
        HM452984  .......... .......... .......... .......... .......... .......... .......... .......... .......... 171
        HM452985  .......... .......... .......... .......... .......... .......... .......... .......... .......... 171
        HM452986  ...P...... ..Q....... .........S ......S... ....PAQLE. .....F.... ...[R].... ........W.. ....R.S... 179
                  Non-repetitive domain (NR1)  |◀──────        Q2        ──────▶|
        HM452979  QQL[CC]QQLRQ IPDQSR[C]QAI HSVVHAIILH QQQLQQQQQQ QQQQQQQQQQ QQ[Q]QQQQQQQY PPSQGSFQQP RQQYPSDQGS FQHPQQQYPS 261
        HM452980  ...|..|.... ......|..|.. .......... .......... .......... .........-. .......... .......... ..Q....... 260
        HM452981  ...|..|..W. ..E...|..|.. ....Q..... H..Q...... .......---- ---------- ----.S.... ..Q......L 219
        HM452982  ...|..|..W. ..E...|..|.. ....Q..... ...Q...... .......--- ---------- ---------- ----.S.... ..Q......L 215
        HM452983  H..|..|.... ......|..|.. .......... .......... .......... .....----. .......... .......... ..Q....... 257
        HM452984  ...|..|.... ......|..|.. .....V.... .......... .......... .........-. .......... .......... ..Q....... 260
        HM452985  ...|..|.... ..E...|..|.. ....Q..... .......... .......... .........-. .......... .......... ..Q....... 260
        HM452986  ..Q|..|..W. ..E...|..|.. ....Q..... ...QL..... .........-- ---------- Q.....S.V. I.Q......L 257
                     Non-repetitive domain (NR2)
        HM452979  G[Q]GSFQPSQQ NPQGQGSAQP QQLPQFEEIR NLALQTLPTM [C]NVYVHPY[C]S TTIAPLGSID TN 323
        HM452980  .|........| ........V. .......... .......... .|......|. .......... .. 322
        HM452981  .|........| ..H..S.V.. ....R..... .I........ .|....P..|. A....F..TG .. 281
        HM452982  .|........| .....S.V.H ....K..... .I........ .|....P..|. A....F..TG .. 277
        HM452983  .|........| ........V. .......... .......... .|........|. .......... .. 319
        HM452984  .|........| ........V. .......... .......... .|........|. .......... .. 322
        HM452985  .|........| ........V. .......... .......... .|........|. .......... .. 322
        HM452986  V|.....E..A| ......FV.. ..VS.I.... R.T..M...L .|....P..|. ..T.....TG V. 319
                  |QGSFQPSQQ|
                   Glia-α
```

Figure 2. Comparison of the deduced amino acid sequences of the 8 α-gliadin full ORFs. The Q1 and Q2 represent the two polyglutamine repeats. The cysteine residues are boxed. The epitopes Glia-α (QGSFQPSQQ) was showed under the alignment
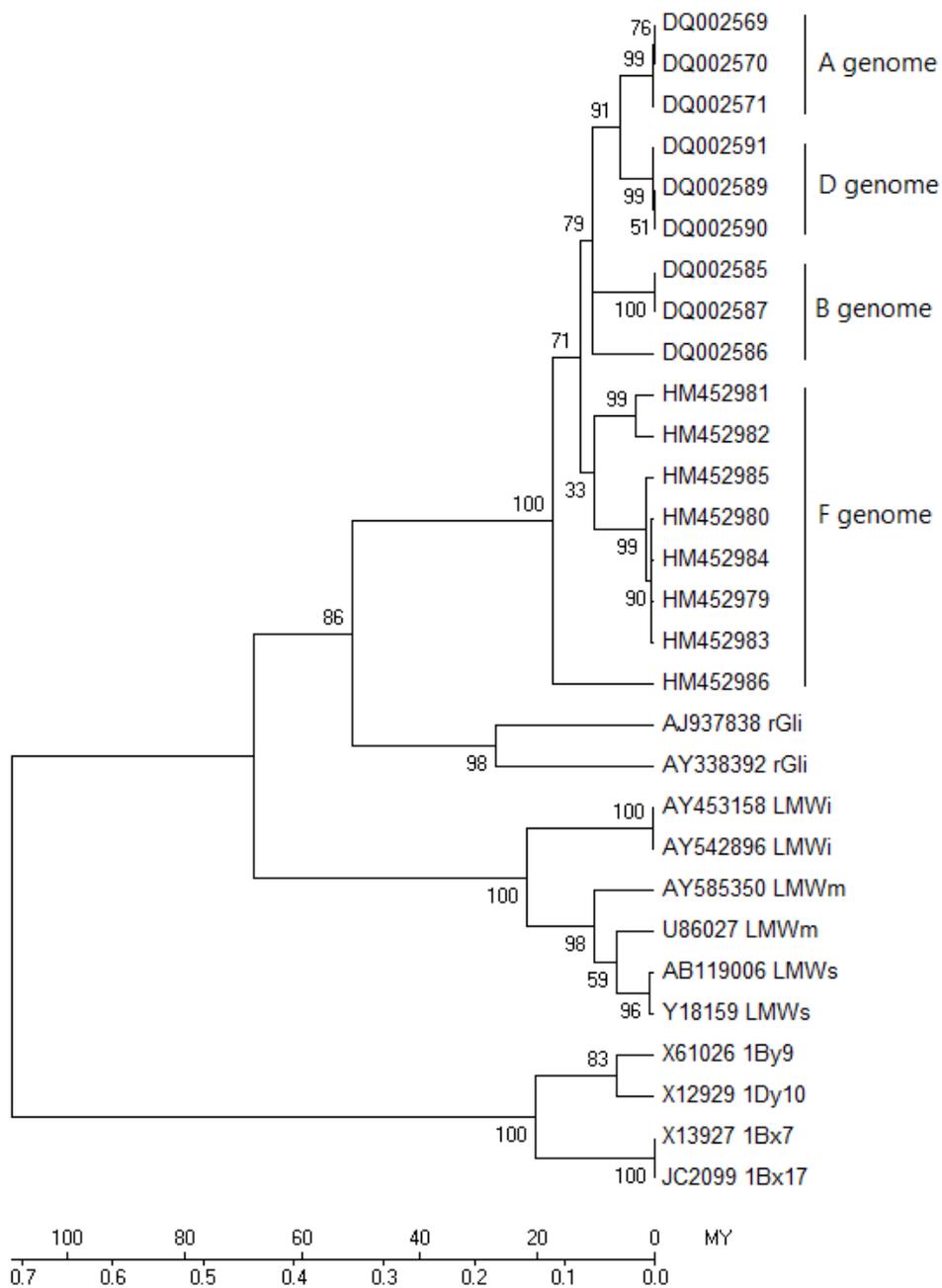
Figure 3. Phylogenetic tree among the *E. bonaepartis* α-gliadin and other storage protein genes from *Triticum* species