Phaseolin nucleotide sequence diversity in *Phaseolus*. I. Intraspecific diversity in *Phaseolus vulgaris*

JAMES A. KAMI AND PAUL GEPTS
Department of Agronomy and Range Science, University of California, Davis, CA 95616-8515, U.S.A.

Corresponding Editor: B. Golding
Received January 6, 1994
Accepted April 26, 1994

KAMI, J.A., and GEPTS, P. 1994. Phaseolin nucleotide sequence diversity in *Phaseolus*. I. Intraspecific diversity in *Phaseolus vulgaris*. Genome, **37**: 751–757.

Most information about the molecular biology of phaseolin, the major seed storage protein in *Phaseolus vulgaris*, has been obtained from the T-type phaseolin, which is characteristic of the Andean gene pool of the species. In the work reported here, two cDNA clones for the S-type phaseolin representing the other major, Middle American gene pool were isolated and sequenced. Analysis of the DNA sequences revealed the presence of two subtypes of S phaseolin, α and β, depending on the presence or absence, respectively, of a 27-bp direct repeat. These are similar to the α- and β-phaseolin subtypes found in the Andean, T phaseolin; however, the additional 15-bp direct repeat also found in the T α-phaseolin gene type was apparently absent from the S α-phaseolin genes. The overall sequence identity was greater between the α or β subtypes of different gene pools than between the α or β subtypes within gene pools. This implies that the gene subtypes were formed prior to the formation of the two major gene pools of *P. vulgaris*. Analysis of the putative amino acid sequence revealed that both the 'Sanilac' phaseolin subtypes contained an additional methionine, however, not at the same site. This opens the possibility of increasing the nutritionally limiting methionine level in phaseolin either through protein engineering or by screening accessions for recombinant phaseolin sequences that combine both substitutions.

Key words: seed storage protein, multigene family, direct repeat.

KAMI, J.A., et GEPTS, P. 1994. Phaseolin nucleotide sequence diversity in *Phaseolus*. I. Intraspecific diversity in *Phaseolus vulgaris*. Genome, **37** : 751–757.

Les informations sur la biologie moléculaire de la phaséoline, la principale protéine de réserve des graines de *Phaseolus vulgaris*, relèvent en général du type T de phaséoline, lequel est caractéristique du pool génique andin des gènes de l'espèce. Dans les travaux qui font l'objet de cet article, deux clones d'ADNc de phaséoline de type S ont été isolés et séquencés et représentent l'autre pool génique majeur, mésoaméricain. L'analyse des séquences d'ADN a révélé la présence de deux sous-types de phaséoline S, α et β, dépendant de la présence ou de l'absence respective d'une répétition directe de 27 pb. Ces sous-types sont similaires aux sous-types andins α et β de la phaséoline T; toutefois, une répétition additionnelle de 15 pb a été trouvée chez le type de gène de l'α-phaséoline T, laquelle est apparemment absente des gènes de l'α-phaséoline S. Dans l'ensemble, l'identité des séquences a été plus grande entre les sous-types α ou β des différents pools géniques qu'entre les sous-types α ou β à l'intérieur des pools géniques. Ceci implique que des sous-types de gènes ont été formés antérieurement à la formation des deux principaux pools géniques de *P. vulgaris*. L'analyse des séquences présumées des acides aminés a révélé que les deux sous-types de phaséoline 'Sanilac' contenaient une méthionine additionnelle, bien que ne provenant pas du même site. Ceci mène à la possibilité d'augmenter le niveau limitatif de la méthionine sur le plan nutritionnel, soit par manipulation génétique, soit par sélection des accessions pour les séquences recombinantes de la phaséoline qui combinent les deux substitutions.

Mots clés : protéine de réserve des graines, famille multigénique, répétition directe.

[Traduit par la rédaction]

Introduction

Phaseolin is the major seed storage protein of *Phaseolus vulgaris* (common bean) and represents up to 49% of the total nitrogen of the seed (Ma and Bliss 1978). It has a sedimentation rate of about 7S and, as such, is placed in the vicilin class of seed storage proteins (Blagrove et al. 1984). Recently, phaseolin protein diversity (as observed by one-dimensional SDS–PAGE) has provided evidence for a subdivision of *P. vulgaris* into two major gene pools (Gepts et al. 1986; Gepts 1988*a*, 1988*b*; Koenig and Gepts 1989; Singh et al. 1991). These gene pools are the Middle American group in which the S-type phaseolin (originally described in the cultivar 'Sanilac') predominates, and the Andean group, which expresses primarily the T-type phaseolin, originally described in 'Tendergreen' (Gepts 1990, 1993; Gepts and Bliss 1986; Gepts et al. 1986; Becerra Velásquez and Gepts 1994). Subsequent work has demonstrated that this distinction can be extended to the nucleic acid level where the S and T types can be distinguished by unique polymorphisms in a Southern hybridization (Talbot et al. 1984; Nodari et al. 1993).

At the present time, only phaseolin genes of the T type of Andean origin have been extensively investigated at the molecular level (Hall et al. 1984; Lawrence et al. 1990; Slightom et al. 1983, 1985; Talbot et al. 1984). Sequencing of phaseolin cDNAs from the Andean cultivar 'Tendergreen' indicated that two phaseolin gene subfamilies, α and β, are, in part, responsible for the complex protein pattern observed in SDS–PAGE. The two gene types differ primarily in that the α-phaseolin genes contain two, small (15 and 27 bp) direct repeats, whereas the β-phaseolin genes do not contain these repeats (Slightom et al. 1985). The presence of these repeats in combination with post-translational glycosylation (Lioi and Bollini 1984; Sturm et al. 1987) is responsible for the distinctive T-phaseolin protein pattern on one-dimensional SDS–PAGE gels.

Printed in Canada / Imprimé au Canada

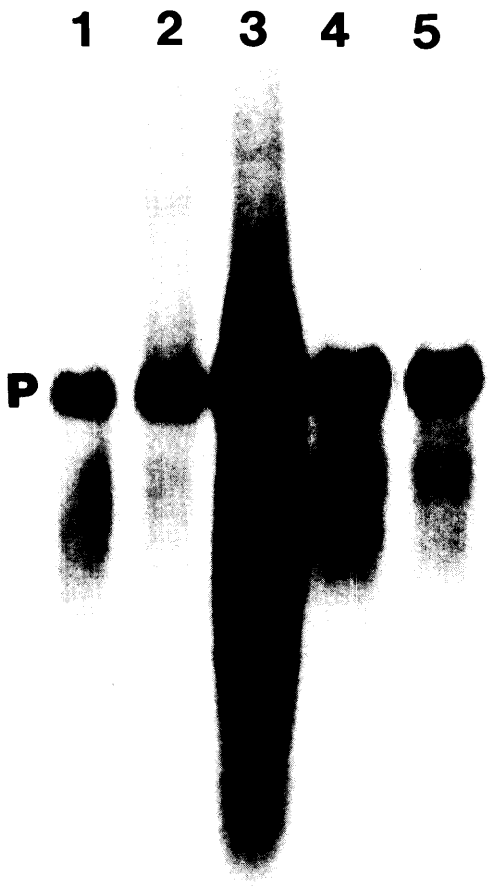

FIG. 1. Northern blot of mRNA from developing cotyledons. Approximately 3 μg poly(A) RNA per sample, probed with a 'Tendergreen' α-phaseolin cDNA clone. Lane 1, *Phaseolus vulgaris* 'Tendergreen' (Andean); lane 2, *Phaseolus vulgaris* 'Contender' Andean); lane 3, *Phaseolus vulgaris* 'Sanilac' (Middle American); lane 4, *Phaseolus lunatus* 'Henderson' (Middle American); lane 5, *Phaseolus lunatus* 'G25804' (Andean); lane 6, *Vigna unguiculata* 'BE85-24' (African). P, position of T phaseolin transcript (1535 bases, without poly(A) tail).

To extend this investigation of the sources of the unique protein polymorphisms that help define the two major gene pools of *P. vulgaris*, additional sequence analysis of the S-type phaseolin genes was needed. In addition, sequence analysis of the phaseolin type that characterizes the other major gene pool of *P. vulgaris* would make it possible to deduce potential sites for amino acid replacement to improve the overall methionine content of phaseolin and thus overcome the primary nutritional deficiency present in legumes (Gepts and Bliss 1984; Ma and Bliss 1978).

Materials and methods

Plant materials

Seeds of the genotypes used for mRNA analysis and cDNA construction and analysis were obtained from the University of California, Davis, bean genetics collection. These included *P. vulgaris* 'Sanilac' and 'Tendergreen', *P. lunatus* 'Henderson' and 'G25804', and *V. unguiculata* 'BE85-24'.

Seed protein extraction

Seed proteins were extracted by the method of Ma and Bliss (1978). SDS–PAGE electrophoresis was performed by the method of Laemmli (1970). Samples, either seed flour or protein fractions, were suspended in a modified SDS sample buffer (0.25 M Tris–HCl (pH 6.8), 0.5 M NaCl, 1% SDS, 1% mercaptoethanol) and heated to 100°C for 10 min before loading.

Plasmid isolation and probe labeling

Plasmids were routinely isolated using the alkaline-SDS method as described in Sambrook et al. (1989). For sequencing and vector preparation, plasmids were further purified by polyethylene glycol precipitation. Probes were labeled with [α-^{32}P]dCTP using the random primer method of Feinberg and Vogelstein (1983).

RNA isolation

Total RNA was isolated essentially by the method of Hall et al. (1978) except that the buffer/seed ratio was increased to 5 mL/g and the Proteinase K concentration increased to 2 mg/g starting material. Poly(A) RNA was isolated by chromatography over oligo(dT)-cellulose according to the method described by Sambrook et al. (1989).

Northern blot hybridization

Northern blotting was performed following the method of Fourney et al. (1988). Following hybridization, washes included 2 × 20 min in 1× SSC, 0.1% SDS at room temperature, 2 × 20 min in 0.1× SSC, 0.1% SDS at 55°C. Blots were then autoradiographed for 12–48 h at −70°C.

cDNA synthesis

cDNA was synthesized with Superscript™ enzyme (Moloney Murine Leukemia Virus, RNase H mint Reverse Transcriptase [Gibco–BRL, Gaithersburg, Md.]) according to the method supplied with the enzyme. After blunt ending the cDNA with T4 polymerase, *Sal*I linkers were ligated to the cDNA. Following overnight digestion with 50 units of *Sal*I enzyme, the oligonucleotides were removed and the cDNA size selected by passage over a Sepharose CL-4B column. The cDNA was recovered by ethanol precipitation and ligated into pUC19 that had been cut with *Sal*I, dephosphorylated with alkaline phosphatase, and gel purified.

Transformation, colony lifts, and hybridization

Escherichia coli cells (MV1190) were transformed the Colony Transformation Protocol of Hanahan (1985). Colony lifts were performed as described in Sambrook et al. (1989). Filters were prehybridized in 6× SSPE, 50% formamide, 0.05× BLOTTO (5% nonfat dry milk in water) at 42°C for 2–3 h, then hybridized in fresh buffer containing 2–5 × 10^5 cpm/mL ^{32}P-labeled probe (pPHAS MC31 α-phaseolin cDNA from 'Tendergreen') for 12–24 h. Washes were 3 × 5 min in 2× SSC, 0.1% SDS at room temperature, and 2 × 9 min in 1× SSC, 0.1% SDS at 68°C. The filters were then blotted dry and autoradiographed at −70°C overnight.

Single-stranded DNA preparation and RD nested deletion series

Single-stranded DNA (ssDNA) for sequencing and for generating nested deletion series was prepared by the method of Vieira and Messing (1987). A nested deletion series for sequencing was constructed by the method of Dale and Arrow (1987).

DNA sequencing

Sequencing of DNA was performed using Sequenase v.2 (U.S. Biochemical, Cleveland, Ohio) according to the manufacturer's instructions.

Computer sequence analyses

Sequence comparisons and alignments were performed with the Wisconsin Genetics Computer Group software version 7.2 (Devereux et al. 1991). The DNA sequences for the S α-phaseolin (clone 2-13) and β-phaseolin (clone 1–12) have been deposited with GenBank (accession numbers U01132 and U01131, respectively).

Results

Isolation and sequence analysis of S phaseolin clones 2-13 and 1-12

Northern blot analysis of total RNA isolated from *P. vulgaris* ('Tendergreen', 'Sanilac', and 'Contender'), *P. lunatus* ('Henderson' and 'G25804'), and *Vigna unguiculata* ('BE85-24'), when hybridized to a 'Tendergreen' α-phaseolin cDNA clone (kindly provided by Dr. J. Slightom, Upjohn, Kalamazoo, Mich.), showed a single, major band of similar molecular size in all five samples (Fig. 1). These results indicate that *Phaseolus* spp. and a species from the related genus *Vigna* contain at least one very highly abundant transcript that has a strong homology to the phaseolin gene and has approximately the same molecular size as the phaseolin transcripts in 'Tendergreen'.

Twenty-eight positive cDNA clones were identified from the 'Sanilac' library. After confirming similarity to the T-type α-phaseolin by sequencing the 3′ end, two clones were selected for complete sequence analysis. Clone 2-13 contained the 27-bp direct repeat and thus resembled a T-type α-phaseolin gene, whereas clone 1-12 lacked the 27-bp repeat and thus resembled a T-type β-phaseolin gene. Analysis of clone 2-13 showed an insert of 1439 bp in the *Sal*I site of pUC19 (Fig. 2). It contained a perfect 27-bp direct repeat at position 1218–1271. In addition, four other smaller, imperfect duplications were found. These repeats were all 11 bp or less and contained at least one mismatch. They were found in all other phaseolin sequences reported, implying that these repeated sequences were formed prior to the expansion of the phaseolin progenitor into a multigene family. The formation of repeats, whether direct or not, appears therefore to have taken place repeatedly during phaseolin evolution. The S phaseolin cDNA clone 1-12 was slighty longer than cDNA clone 2-13 with an insert size of 1483 bp but did not contain either the 15-bp or the 27-bp direct repeats (Fig. 2). Comparison of the S phaseolin cDNA clones with the 'Tendergreen' cDNA clones (Slightom et al. 1985) indicated that the S α-phaseolin clone 2-13 contained the entire peptide coding region but lacked any of the 5′ noncoding sequences (Fig. 2). The S β-phaseolin clone 1-12 contained a portion of the 5′ noncoding region but did not appear to be a full length cDNA sequence. A comparison between the coding regions of S phaseolin clones 2-13 and 1-12 showed 98.86% sequence identity between the two clones (Table 1). A total of 16, dispersed, single-base substitutions and three gaps were identified. Of the three gaps, one was accounted for by the presence of the 27-bp repeat in clone 2-13 and its absence in clone 1-12. The other two were single-base deletions in the 3′ region of clone 1-12. No 15-bp repeat was observed in the two S phaseolin clones, confirming an earlier observation by Anthony et al. (1990). A polymerase chain reaction test of total genomic DNA of 'Sanilac' with primers bracketing the 15-bp direct repeat region also failed to reveal the presence of this direct repeat in other, uncloned sequences (J. Kami and P. Gepts, unpublished results).

A comparison of sequences from different phaseolin types (e.g., S and T) should be made with caution, as at this stage it is not possible to distinguish orthologous from paralogous members of the respective gene families. With this note of caution in mind, these comparisons show that α-phaseolin sequences of the two phaseolin types (S and T) were more similar to each other than to their respective β-phaseolin sequences. The same statement can be made for the β-phaseolin sequences of the S and T phaseolins (Table 1). This suggested that the divergence into α- and β-phaseolins predates the divergence of the Middle America (S) and Andean (T) phaseolins.

TABLE 1. Percent DNA sequence identity between phaseolin gene types in *Phaseolus vulgaris*

	T-β	S-α	S-β
T-α	98.98% (14) [6]	99.72% (4) [4]	98.86% (16) [9]
T-β		98.89% (15) [8]	99.38% (9) [4]
S-α			98.86% (16) [8]

NOTE: Number of nucleotide substitutions and the number of amino acid replacements are indicated in parentheses and brackets, respectively. T-α and T-β were from Slightom et al. (1985). The sequence of T-α clone 169 was used for comparison. For S-α phaseolin 'Sanilac' clone 2-13 was used. For S-β phaseolin 'Sanilac' clone 1-12 was used.

Phaseolin-derived peptide composition and sequence

Translation of the open reading frame of S α-phaseolin clone 2-13 yielded a polypeptide of 430 amino acids with a calculated molecular mass of 49 270 Da. S β-phaseolin clone 1-12 coded for a polypeptide of 47 520 Da with 422 amino acid residues. When the signal sequence was removed, the molecular masses dropped to 45 904 and 44 995 Da, respectively. The predicted amino acid sequence showed that both the α and β S phaseolin proteins contain an additional methionine residue relative to the T phaseolin proteins confirming the observation of Anthony et al. (1990). Alignment of the peptide sequences showed that the sites of amino acid replacements were more or less randomly distributed throughout the entire sequence (Kami 1993). In general, as with the DNA sequences, there were fewer replacements between the T and S α-phaseolins and between the T and S β-phaseolins than between the α- and β-phaseolin within each cultivar. Of the 12 sites where amino acid replacements had taken place, 8 (positions 2, 9, 170, 208, 245, 274, 375, and 399) involved single replacements in individual genes: 3 in S α-phaseolin clone 2-13, 3 in 'Sanilac' β-phaseolin clone 1-12, and 1 each in T-phaseolin α and β. However, two of the replacements in 'Sanilac' β-phaseolin occurred in the signal sequence region and therefore did not directly affect the mature phaseolin protein. The other four (positions 231, 317, 393, and 410) involved identical replacements in two genes. In each case, the difference was between the T-α/S 2-13 and T-β/S 1-12 pairs. The mutations responsible for these double replacements must have occurred after the initial duplication of the original gene and formation of the 27-bp repeat but before the divergence of the S and T phaseolins and subsequent introduction of the 15-bp repeat. Most of the amino acid replacements were of a highly conservative nature, i.e., their approximate size, shape, charge, and helix-inducing properties were largely conserved by the replacement, implying that they would do little to perturb the three-dimensional structure of the final protein.

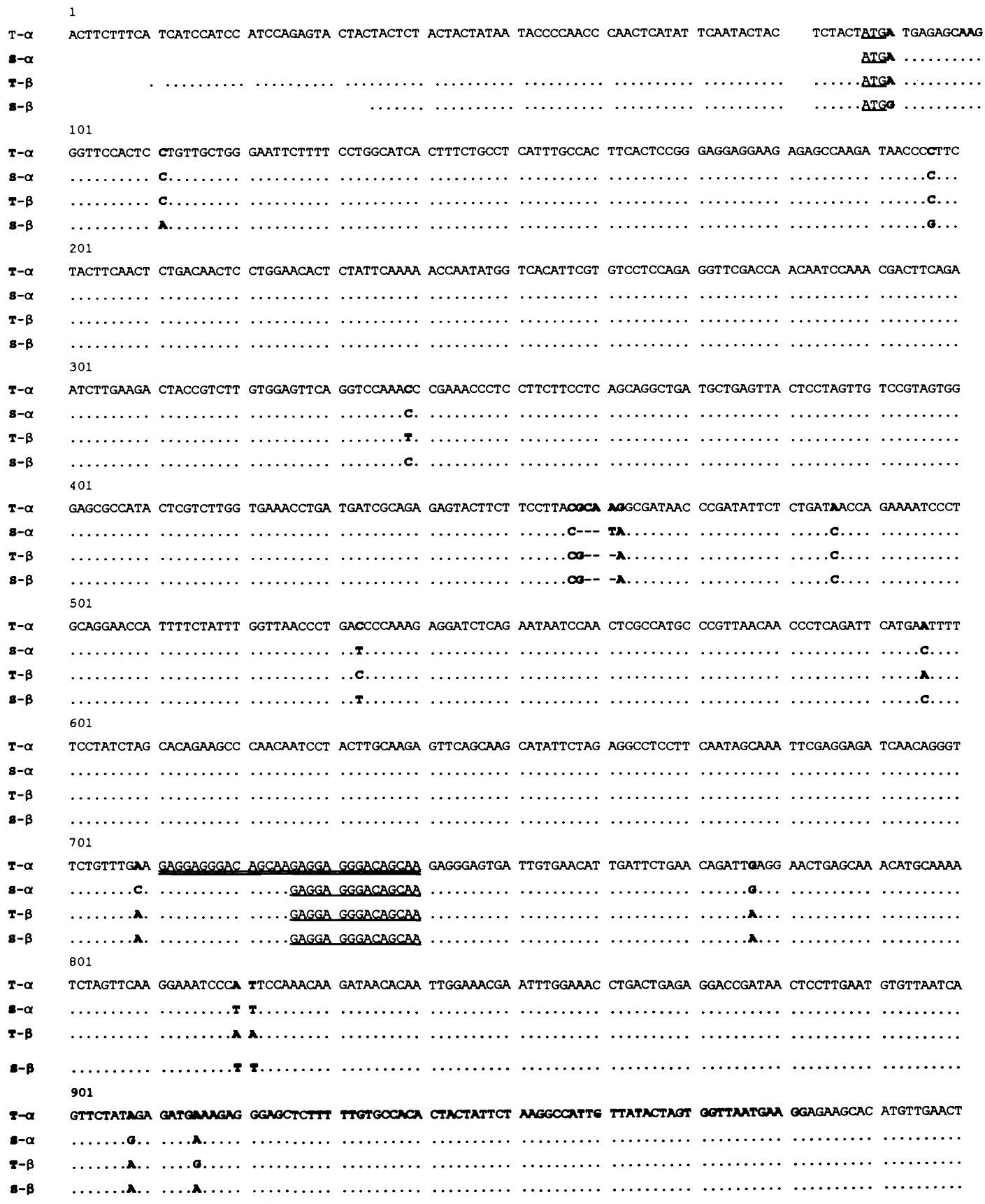

FIG. 2. Comparison of the nucleic acid sequences from the phaseolin cDNA clones. Bold type indicates nucleotide substitutions. Nucleotide sequences for the T α-phaseolin and β-phaseolin nucleotide sequences are from Slightom et al. (1985). The S α-phaseolin and β-phaseolin nucleotide sequences are from clones 2-13 and 1-12, respectively (present results).

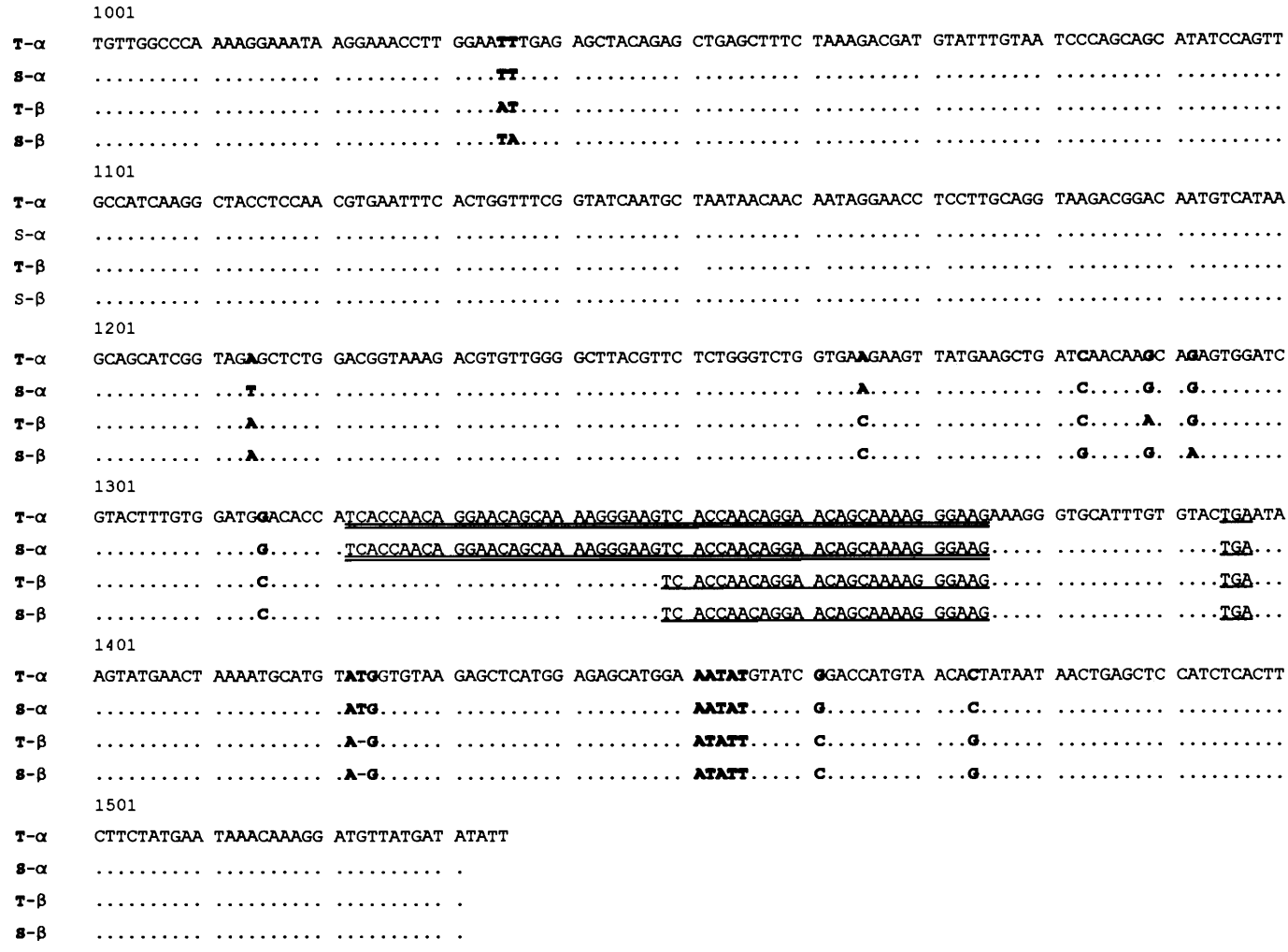

FIG. 2. *Concluded.*

Discussion

Our observations on the nucleotide sequence of the S-type phaseolin genes, in comparison with that of those from T-phaseolin sequences, allow us to sketch a preliminary sequence of events characterizing the molecular evolution of this multigene family in *P. vulgaris*. Given that it is more plausible that the repeats were generated from preexisting sequences, the simplest sequences, that of the β-phaseolins, which lack both the 15- and 27-bp repeats, are most likely the progenitors of the α-phaseolin genes. Since this progenitor sequence is present in both the S and T multigene families along with modified copies, i.e., the repeat-containing α-phaseolins, the β-phaseolin genes would have undergone a duplication event followed by the introduction of repeated sequences into one of the duplicates. Because the 27-bp direct repeat is present in both S and T phaseolins, its introduction took place presumably before the divergence of *P. vulgaris* into its two major geographic, Middle American and Andean, gene pools (Gepts 1988*b*, 1993).

After this geographical separation, a 15-bp repeat was then introduced among Andean α-phaseolin sequences. Concurrently with these events, additional duplications of the phaseolin genes may have taken place, further expanding the gene family. The precise timing of these expansion events is difficult to determine. Since there are an estimated 7–10 members of the gene family, and at least six T α-phaseolin variants have been identified (Slightom et al. 1985), it appears that this expansion may have taken place at least partly after the Andean and Middle American gene pools diverged. However, a similar number of genes has been estimated for 'Sanilac' (Slightom et al. 1985). This implies that the expansion of the gene family occurred both in the Andean and Middle American gene pools to the same degree. Whether this expansion of the gene families is an apparently random, nonselective event or the result of the selection pressures placed upon genotypes prior, during, or after domestication is not apparent from the present data. However, it should be noted that the S and T protein patterns have been observed in wild-growing beans (Gepts et al. 1986). Unless the S and T phaseolin types were introduced into wild beans through occasional outcrosses with cultivated material, their existence in wild beans suggest that the expansion of the phaseolin multigene family may have taken place prior to domestication.

Thus, the evolution of the phaseolin genes in the two primary gene pools of *P. vulgaris* can be seen to be punctuated by a series of major events: (*i*) the duplication of the progenitor sequence into at least two genes to start the multiplication of the multigene family; (*ii*) the generation of a 27-bp direct repeat in one of the duplicates of the original gene; (*iii*) the acquisition of at least four mutations within the coding region of the genes; (*iv*) the geographical separation

of the S and T phaseolin types; (*v*) the introduction of the 15-bp repeat into the same gene as the 27-bp repeat in the T phaseolin type; and (*vi*) concurrently with the previous event there was further expansion of the multigene family by gene duplication.

Since it has been demonstrated that the nutritional availability of methionine in beans is positively correlated with phaseolin protein content (Gepts and Bliss 1984), the S phaseolin can provide a higher nutritional value than the T, other factors being equal. Second, in addition to the increased number of methionine residues in S phaseolin (Anthony et al. 1990 and present results), the smaller S α-phaseolin protein provides a higher molar ratio of methionine than its larger T phaseolin counterpart and should perhaps constitute the starting point for sequence modifications to enhance phaseolin methionine content.

A comparison of sequences of different phaseolin types should allow us to reveal additional amino acid sites that are variable and thus potentially amenable to sequence modification to improve the amino acid balance of phaseolin. This is especially true as sequence homogenization has been posited for the phaseolin multigene family (Slightom et al. 1985). Although no extended regions of nonconserved amino acid sequences could be identified that would be preferential targets for sequence modification, single amino acid replacements appear to be tolerated at specific sites. Hence, the possible introduction of additional methionine residues into the protein through gene modification at specific sites does appear to be feasible. The replacement of I to M and M to I at positions 274 and 399, respectively, in both the S α- and β-phaseolin peptides has possible important nutritional implications. Since these proteins seem to tolerate the replacements, there exists the potential to modify a phaseolin gene in vitro to contain methionine residues at both positions within the same coding sequence. Alternatively, different accessions of *P. vulgaris* could be screened to identify a possible recombinant that contains both methionines in the same sequence. This would increase the overall methionine content by 20% and the mole percent to about 1.23. Although not a complete solution to the problem of methionine as the nutritionally limiting amino acid, it would be an improvement on present varieties.

Acknowledgments

Funding for this research was provided by the International Board for Plant Genetic Resources, Rome. We thank Dr. J. Slightom for providing clones of T-phaseolin sequences.

Anthony, J.L., Vonder Haar, R.A., and Hall, T.C. 1990. Nucleotide sequence of an α-phaseolin gene from *Phaseolus vulgaris*. Nucleic Acids Res. **18**: 3396.

Becerra Velásquez, V.L., and Gepts, P. 1994. RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origins. Genome, **37**: 256–263.

Blagrove, R.J., Lilley, G.G., Donkelaar, A.V., Sun, S.M., and Hall, T.C. 1984. Structural studies of a French bean storage protein: phaseolin. Int. J. Macromol. **6**: 137–141.

Dale, R.M.K., and Arrow, A. 1987. A rapid single-stranded cloning, sequencing, insertion and deletion stratagy. Methods Enzymol. **155**: 204–214.

Devereux, J., Haeberli, P., and Smithies, O. 1991. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**: 387–395.

Feinberg, A.P., and Vogelstein 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**: 9–13.

Fourney, R.M., Miyakoshi, J., Day, R.S., and Paterson, M.C. 1988. Northern blotting: efficient RNA staining and transfer. Focus, **10**: 5–7.

Gepts, P. 1988*a*. Phaseolin as an evolutionary marker. *In* Genetic resources of *Phaseolus* beans. *Edited by* P. Gepts. Kluwer, Dordrecht, The Netherlands. pp. 215–241.

Gepts, P. 1988*b*. A Middle American and Andean gene pool. *In* Genetic resources of *Phaseolus* beans. *Edited by* P. Gepts. Kluwer, Dordrecht. The Netherlands. pp. 375–390.

Gepts, P. 1990. Biochemical evidence bearing on the domestication of *Phaseolus* beans. Econ. Bot. **44**(3S): 28–38.

Gepts, P. 1993. The use of molecular and biochemical markers in crop evolution studies. Evol. Biol. **27**: 51–94.

Gepts, R., and Bliss, F.A. 1984. Enhanced available methionine concentration associated with higher phaseolin levels in common bean seeds. Theor. Appl. Genet. **69**: 47–53.

Gepts, P., and Bliss, F.A. 1986. Phaseolin variability among wild and cultivated common beans (*Phaseolus vulgaris*) from Colombia. Econ. Bot. **40**: 469–478.

Gepts, P., Osborn, T.C., Rashka, K., and Bliss, F.A. 1986. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): evidence for multiple centers of domestication. Econ. Bot. **40**: 451–468.

Hall, T.C., Ma, Y., Buchbinder, B.U., Pyne, J.W., Sun, S.M., and Bliss, F.A. 1978. Messenger RNA for G1 protein of French bean seeds: cell-free translation and product characterization. Proc. Natl. Acad. Sci. U.S.A. **75**: 3196–3200.

Hall, T.C., Slightom, J.L., Ersland, D.R., Scharf, P., Barker, R.F., Murray, M.G., Brown, J.W.S., and Kemp, J.D. 1984. The phaseolin family of seed protein genes: sequences and properties. *In* Advances in gene technology: molecular genetics of plants and animals. *Edited by* K. Downey, J. Schultz, F. Ahmad, and R.W. Voellnig. Academic Press, New York. pp. 349–367.

Hanahan, D. 1985. Techniques for transformation of *E. coli*. *In* DNA cloning: a practical approach. *Edited by* D.M. Glover. IRL, Oxford, U.K. pp. 109–135.

Kami, J. 1993. The molecular evolution of phaseolin. Ph.D. dissertation, University of California, Davis, Calif.

Koenig, R., and Gepts, P. 1989. Allozyme diversity in wild *Phaseolus vulgaris*: further evidence for two major centers of diversity. Theor. Appl. Genet. **78**: 809–817.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London), **227**: 680–685.

Lawrence, M.C., Suzuki, E., Varghese, J.N., Davis, P.C., Van Donkelaar, A., Tulloch, P.A., and Colman, P.M. 1990. The three-dimensional structure of the seed storage protein phaseolin at 3 Å resolution. EMBO J. **9**: 9–15.

Lioi, L., and Bollini, R. 1984. Contribution of processing events to the molecular heterogeneity of four banding types of phaseolin, the major storage protein of *Phaseolus vulgaris* L. Plant Mol. Biol. **3**: 345–353.

Ma, Y., and Bliss, F.A. 1978. Seed proteins of common bean. Crop Sci. **17**: 431–437.

Nodari, R.O., Tsai, S.M., Gilbertson, R.L., and Gepts, P. 1993. Towards an integrated linkage map of common bean. II. Development of an RFLP-based linkage map. Theor. Appl. Genet. **85**: 513–520.

Sambrook, J., Fritsch, E., and Maniatis, T. 1989. Molecular Cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Singh, S.P., Gutiérrez, J.A., Molina, A., Urrea, C., and Gepts, P. 1991. Genetic diversity in cultivated common bean. II. Marker-based analysis of morphological and agronomic traits. Crop Sci. **31**: 23–29.

Slightom, J.L., Sun, S.M., and Hall, T.C. 1983. Complete

nucleotide sequence of a French bean storage protein gene: phaseolin. Proc. Natl. Acad. Sci. U.S.A. **80**: 1897–1901.

Slightom, J.L., Drong, R.F., Klassy, R.C., and Hoffman, L.M. 1985. Nucleotide sequences from phaseolin cDNA clones: the major storage proteins from *Phaseolus vulgaris* are encoded by two unique gene families. Nucleic Acids Res. **13**: 6483–6498.

Sturm, A., Van Kuik, J.A., Vliegenthart, J.F.G., and Chrispeels, M.J. 1987. Structure, position, and cDNA biosynthesis of the high-mannose and the complex oligosaccharide side chains of the bean storage protein phaseolin. J. Biol. Chem. **262**: 13392–13403.

Talbot, D.R., Adang, M.J., Slightom, J.L., and Hall, T.C. 1984. Size and organization of a multigene family encoding phaseolin, the major seed storage protein of *Phaseolus vulgaris* L. Mol. Gen. Genet. **198**: 42–49.

Vieira, J., and Messing, J. 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**: 3–18.