Phaseolin has proven to be an excellent marker to uncover some basic features of the genetic diversity in common bean, Phaseolus vulgaris: 1) the reduction of diversity during domestication (Gepts and Bliss 1986; Gepts et al. 1986; Koenig et al. 1990); 2) the existence of two major subdivisions ("subspecies") with a distinct geographic distribution, Mesoamerica vs. Andes (Gepts et al. 1986); and 3) the identification of certain races, such as races Durango and Chile (Singh et al. 1991).

The underlying molecular mechanisms responsible for phaseolin diversity are poorly known. This information could, however, be useful to determine phylogenetic relationships among the various segments of Phaseolus germplasm and possibly to identify non-conserved regions of the phaseolin gene that would be more amenable to sequence modification for the purpose of improving the nutritional qualities of phaseolin. To obtain this information, we have embarked on a project that aims at sequencing phaseolin genes from Mesoamerican Phaseolus vulgaris, from Mesoamerican and Andean Phaseolus lunatus, and from Vigna unguiculata. (The DNA sequence of an Andean phaseolin - 'T phaseolin has already been determined by Sun et al. 1981 and Slightom et al. 1985.) This approach will allow us to compare DNA sequence diversification at the intergeneric, interspecific, and intraspecific levels.

In a first stage of this project, we have isolated and sequenced a phaseolin cDNA clone from the Mesoamerican Phaseolus vulgaris cultivar 'Sanilac'. The clone was isolated from a cDNA library prepared from RNA purified from developing cotyledons. The clone was identified with a radiolabeled probe prepared from a 'T type phaseolin cDNA clone, MC31 (Slightom et al. 1983).

Comparison of the 1446 bp Sanilac ('S' type) clone with the full length 'T type clone revealed greater than 99% sequence homology. The Sanilac clone begins at position nt 93 of the 'T' phaseolin clone, thus is missing the leader sequence and the first amino acid codon. The most striking feature about the clone is the presence of a 27 bp direct repeat that matches exactly to a similar repeat in the 'T' type clone. However, a 15 bp repeat, also in the 'T' type clone is missing from the 'S' type clone. To date, cDNA clones isolated from 'T' type plants have shown either the presence of both repeats or the absence of both. This is the first time a cDNA has been sequenced that has only one of the repeats. This correlates with a genomic sequence for a Sanilac phaseolin gene previously reported (Anthony et al. 1990).

An additional correlation to the genomic sequence is the presence of a methionine codon which has been substituted for an isoleucine codon in the 'T' type sequence. From a protein engineering standpoint, this is a relatively benign substitution in that methionine and isoleucine are both uncharged amino acids with similar shapes and sizes. Thus this substitution will do little to disturb the three-dimensional structure of the protein.

While the 'S' type sequence has strong homology to the 'T'-type alpha-6 clone (the clone found in Genbank Acc. # X02980), a slightly better fit was found when compared to
clone alpha-169, one of the variant alpha-phaseolin types isolated by Slightom et al. (1985). If these variant types represent individual members of the phaseolin multi-gene family, this implies that the individual genes in this family duplicated and diverged prior to the separation of the Andean ('T' type) and the Mesoamerican ('S' type) Phaseolus vulgaris.


Gepts P, Bliss FA (1986) Phaseolin variability among wild and cultivated common beans (Phaseolus vulgaris) from Colombia. Econ Bot 40:469-78


