Pulsed-field gel electrophoresis analysis of the phaseolin locus region in *Phaseolus vulgaris*

Víctor Llaca and Paul Gepts

Abstract: Phaseolin is the major seed storage protein of common bean (*Phaseolus vulgaris* L.). It is encoded by a small multigene family of 6–9 genes that are clustered in a single complex locus (*Phs*). We have constructed a long-range restriction map of the phaseolin genomic region, including the *Phs* locus and two flanking marker loci, *D1861* and *Bng060*. Using a combination of high molecular weight DNA isolation, one- and two-dimensional pulsed-field gel electrophoresis of single and double restriction digests followed by Southern hybridization, and PCR analysis of individual fragments, we found that: (*i*) the maximum size of the *Phs* locus is 190 kb, (*ii*) the *Phs* locus may have increased in size during the evolution of *P. vulgaris*, (*iii*) the genomic region marked by *D1861*–*Phs*–*Bng060* spans 5 cM, which corresponds to a maximum of 1.9 Mb, and (*iv*) the *Phs* locus could be oriented with respect to the two adjacent markers. Further progress in determining the gene arrangement in the *Phs* locus will require cloning and analysis of large DNA fragments containing phaseolin genes via BAC libraries.

Key words: multigene family, physical distance, genome mapping, seed protein.

Résumé : La phaséoline est la principale protéine de réserve chez le haricot (*Phaseolus vulgaris* L.). Elle est codée par une petite famille de multigènes qui compte 6–9 gènes tous groupés en un seul locus complexe (*Phs*). Une carte de restriction à longue distance de la région du locus de la phaséoline a été produite. Cette région comprend le locus *Phs* ainsi que deux marqueurs adjacents, *D1861* et *Bng060*. Pour réaliser cette analyse, les auteurs ont employé une combinaison de diverses techniques et approches dont : l'isolation d'ADN de haut poids moléculaire, l'électrophorèse en champs pulsés uni- et bi-dimensionnels de digestions enzymatiques simples et doubles suivies d'hybridations Southern et, finalement, l'analyse de fragments individuels par amplification PCR. Ces travaux ont permis de montrer : (*i*) que la taille maximale du locus *Phs* est de 190 kb, (*ii*) que la taille du locus *Phs* peut s'être accrue au cours de l'évolution du *P. vulgaris*, (*iii*) que la région qui comprend *D1861*–*Phs*–*Bng060* mesure 5 cM ce qui correspond tout au plus à 1,9 Mb et (*iv*) que le locus *Phs* a pu être orienté par rapport aux deux marqueurs adjacents. Des études plus détaillées de l'organisation des gènes au niveau du locus *Phs* exigeront le clonage et l'analyse de grands fragments d'ADN contenant les gènes de la phaséoline grâce à des banques BAC.

Mots clés : famille de multigènes, distance physique, cartographie des génomes, protéines de réserve.

[Traduit par la Rédaction]

Introduction

Phaseolin is the major seed storage protein of common bean, as it accounts for 30–50% of total seed nitrogen (Ma and Bliss 1978). A substantial amount of information has been gained at the molecular level about the organization and expression of phaseolin, based principally on analyses of the "T" phaseolin (Bustos et al. 1989; Casey and Domoney 1987; Chee et al. 1986; Cramer et al. 1985, 1987; Murray and Kennard 1984; Sengupta et al. 1985; van Santen and Spritz 1987). According to reassociation kinetics experiments,

genomic and cDNA clone analysis, and Southern blotting, phaseolin is encoded by a small multigene family of 6–9 members (Talbot et al. 1984). Two-dimensional electrophoresis resolved 5–8 different polypeptides (Hall et al. 1984; Talbot et al. 1984). Phaseolin genes are approximately 2000 base pairs (bp) long and contain 6 exons and 5 introns (Anthony et al. 1990; Slightom et al. 1983, 1985). Two subfamilies, α and β, have been identified on the basis of the presence of tandem direct repeats. In *Phaseolus vulgaris* cv. Tendergreen, which carries the "T" phaseolin, the major difference between the α and β phaseolin type genes is the presence or absence, respectively, of three direct repeats. The direct repeats are 21, 15, and 27 bp long and are located in the third intron and the fourth and sixth exons, respectively. In *P. vulgaris* cv. Sanilac, which carries the "S" phaseolin, the α and β phaseolin type genes contain or lack, respectively, the 21- and 27-bp tandem direct repeats. "T" phaseolin genes, the presumed ancestral phaseolin genes found in *P. vulgaris* accession G21245, are all β-phaseolin genes, as they lack any of the tandem direct repeats (Kami et al. 1995).

V. Llaca[1] **and P. Gepts.**[2] Department of Agronomy and Range Science, University of California, Davis, CA 95616-8515, U.S.A.

[1] Present address: Waksman Institute, Rutgers University, Piscataway, NJ 08855-0759, U.S.A.
[2] Author to whom all correspondence should be addressed.

Little is known about aspects related to the chromosomal organization and size of the phaseolin locus (*Phs*). All phaseolin-related sequences are tightly linked genetically in a single complex locus (*Phs*) on linkage group D7; recombination has not been observed among phaseolin genes in crosses between varieties (Brown et al. 1981; Bliss and Brown 1982; Koenig and Gepts 1989; Nodari et al. 1993*a*). Phaseolin has been mapped by in situ hybridization to a single site on chromosome 3 in *Phaseolus coccineus* (Schumann et al. 1990). However, the various members of the gene family are physically separated by at least 13–20 kilobases (kb), as lambda phage genomic libraries have not yielded clones with two or more phaseolin sequences (Hall et al. 1984).

The goal of the research reported here, was to gain additional information on the organization of the phaseolin locus. Specifically, we wanted to determine the maximum size of the locus and how it has evolved during the evolution of the species. We also wanted to determine the relationship between genetic and physical distance in a genomic region marked by *Phs* and two flanking markers and to orient *Phs* with respect to these two loci.

Materials and methods

Plant material
Phaseolus vulgaris cv. Tendergreen, derived from Andean domesticates and with a "T" type phaseolin pattern, was used for most of the pulsed-field gel electrophoresis (PFGE) analysis, because most of the information regarding phaseolin genes at the cDNA and genomic levels has been gathered in this variety (e.g., Talbot et al. 1984; Slightom et al. 1985). An F_2 population, generated from the cross between the Mesoamerican breeding line BAT93 and the Andean landrace Jalo EEP558 (Nodari et al. 1992), was used to map markers around *Phs*. This population was chosen for its high level of polymorphism (Nodari et al. 1992).

Segregation analysis for genetic mapping
Genetic mapping of restriction fragment length polymorphisms (RFLPs) linked to *Phs* was performed according to Nodari et al. (1993*a*). Markers were derived from three independently established maps (Adam-Blondon et al. 1994; Nodari et al. 1993*a*; Vallejos et al. 1992). Membranes containing restriction enzyme digested DNA from individual plants of the F_2 population were subjected to Southern hybridization with the marker probe, as described previously (Nodari et al. 1993*a*). Segregation analysis was performed with the program MAPMAKER version 1.9 (Lander et al. 1987) as described in Nodari et al. (1993*a*).

High molecular weight DNA isolation and restriction digestion
High molecular weight DNA was extracted from etiolated leaves according to the method of Creusot et al. (1992), although the extraction buffer was modified (10 mM sodium citrate, 50 mM Na$_2$EDTA, 600 mM mannitol, 0.1% (w/v) Triton X-100, 0.1% (w/v) BSA, 0.5% β-mercaptoethanol, pH 6.0). Restriction endonuclease digestion for DNA samples imbibed in agarose plugs was performed according to Ganal and Tanksley (1989) and Creusot et al. (1992).

Pulsed-field electrophoresis using contour-clamped homogeneous electric field (CHEF) two-dimensional electrophoresis and Southern hybridization
Pulsed-field gel electrophoresis was carried out according to published protocols (Anand and Southern 1990) using a

CHEF-DRII Unit (Bio-Rad). One percent PFGE-grade agarose (United States Biochemicals) in 0.5× TBE (45 mM Tris-borate, 1 mM EDTA, pH 8.0) was usually used as the matrix. The gel was run at 14°C. Switching times and total running times for separating specific ranges of fragment sizes were set according to van Daelen and Zabel (1992). For routine separations (200–2000 kb), we used a two-stage procedure, with 60 s as the initial switching time for 15 h, followed by 9 h at 90 s at 200 V. For 40–600-kb optimal resolution, we used a one-stage procedure, with a ramped pulse of from 20 to 50 s over 26 h at 200 V. The molecular weight standards were yeast chromosomes (long range), lambda concatamers (long–medium range), and lambda concatamers digested with *Xba*I (short–medium range). Southern hybridization was performed as described previously (Sambrook et al. 1989).

Probes and labeling
For the detection of the phaseolin gene family, two different probes were used, both obtained from the same α-phaseolin cDNA isolated from *P. vulgaris* cv. Sanilac (Kami and Gepts 1994). The first probe was a 1500-kb *Sal*I fragment, containing the whole cDNA fragment, isolated from its pUC19 vector. The second was a 900-bp PCR product obtained by combining the upstream primer for the 15-bp tandem direct repeat PCR test with the downstream primer for the 27-bp tandem direct repeat PCR test (see below and Kami et al. 1995). For the detection of markers flanking *Phs*, *D1861* and *Bng060*, inserts were isolated by digestion of the vector with *Pst*I, separated by low melting point agarose electrophoresis, and excised from the gel. After equilibration in β-agarose buffer (United States Biochemicals), inserts were subjected to β-Agarase digestion, following the protocol recommended by the manufacturer (United States Biochemicals). The random priming method (Feinberg and Vogelstein 1983) was used for labeling DNA fragments (20–30 ng) to be used as probes.

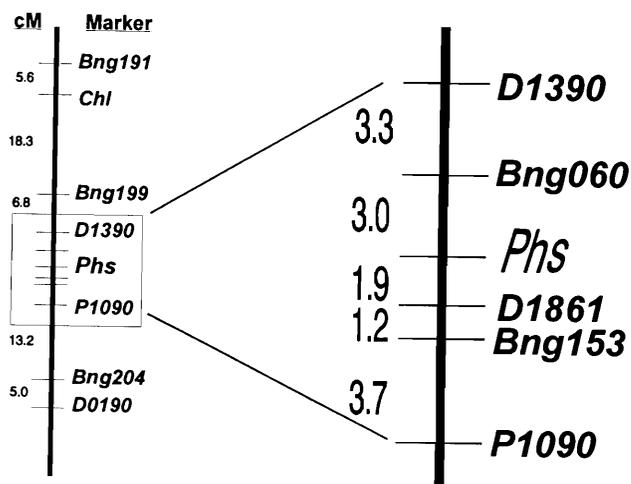### PCR characterization of excised fragments
Fragments of interest were recovered by excising them from low melting point agarose gels at the distance determined by preparative Southern hybridization. Excised fragments were washed several times in TE (10 mM Tris–HCl, 1 mM EDTA, pH 8.0), then equilibrated in β-Agarase I buffer (United States Biochemicals). Fragments were melted at 65°C, transferred to a 37°C water bath, and incubated for 6–7 h in the presence of β-Agarase I according to the manufacturer's protocol (United States Biochemicals). The digested gel sample was extracted once with phenol and once with chloroform – isoamyl alcohol 24:1. It was then precipitated with ethanol, and resuspended in 10 μL TE buffer. DNA concentration was estimated with a dedicated fluorometer. DNA (10 ng) was subjected to PCR as described by Kami et al. (1995). Two sets of primers were used. The first set amplified the third intron and part of the fourth exon, containing the 21- and 15-bp tandem direct repeats, respectively. The second set amplified part of the sixth exon, containing the 27-bp tandem direct repeat (Kami et al. 1995).

Results and discussion

Fine mapping of the phaseolin genomic region
RFLP markers from three sources (Adam-Blondon et al. 1994; Nodari et al. 1992; Vallejos et al. 1992) were mapped in the same population described by Nodari et al. (1992, 1993*a*), consisting of 75 F_2 plants derived from the cross BAT93 × Jalo EEP558. The two nearest markers flanking *Phs* were *D1861* and *Bng060*, located at 1.9 and 3 cM, respectively (Fig. 1).

Fig. 1. Phaseolin genomic region on linkage group D7. Genetic distances are in centimorgans. Linkage was calculated from segregation data obtained in the BAT93 × Jalo EEP558 F_2 segregating population as described in Nodari et al. (1992, 1993a). The map includes RFLPs from three different genomic libraries (D markers, Gepts et al. 1993; Bng markers, Vallejos et al. 1992; P marker, Adam-Blondon et al. 1994) and the chalcone isomerase gene (*Chl*, originally isolated by Mehdy and Lamb 1987).

Table 1. Size of fragments hybridizing to a phaseolin probe after restriction digestion with several enzymes.

Restriction enzyme	Fragment size (kb)
*Bsa*HI	250
*Bss*HII	720
*Eag*I	680
*Mlu*I	350
*Not*I	>2000
*Nru*I	400
*Nsp*V	150+25
*Pin*AI	>200[a]
*Psp*1406	>200
*Sac*II	350
*Sal*I	190
*Sfi*I	>2000
*Sma*I	440
*Sna*BI	>800

[a]The symbol ">" before an estimate means that the size of the fragment was larger than the resolution limit with the CHEF parameters used.

Size of the phaseolin locus

Digestions of high molecular weight genomic DNA from *P. vulgaris* cv. Tendergreen were performed using rare-cutting restriction enzymes, comprising 8-bp cutters and methylation-sensitive 6-bp cutters with a CpG or CpXpG sequence in the recognition site (Siedler and Graner 1991). Table 1 lists the endonucleases producing single fragments hybridizing to the phaseolin probe and the sizes of the fragments. *Sal*I digestions gave the smallest single phaseolin fragment of about 190 kb. This result suggested that all the phaseolin genes were contained in a genomic region of at most 190 kb. This is consistent with the estimates of the number of genes and minimum separation between genes observed in *P. vulgaris* cv. Tendergreen (Hall et al. 1984; Slightom et al. 1983, 1985). The long range restriction map (see next section) provides further evidence for a single unduplicated *Sal*I fragment.

Enzymes recognizing single fragments in the range of 300–500 kb were further used in double digestions to generate a restriction map. These enzymes included *Mlu*I (300 kb), *Sma*I (450 kb), and *Sac*II (350 kb). Neither *Sma*I nor *Sac*II appeared to cut in the *Sal*I fragment containing *Phs*. *Sal*I restriction sites were further mapped using the *Bgl*I endonuclease (Fig. 2). *Bgl*I recognized 5–6 sites internal to *Phs* (not shown). Only one of these *Bgl*I sites was mapped, because it provided information on the orientation of one of the outer genes of the phaseolin gene array. One of the two external *Bgl*I fragments recognized by the phaseolin probe extended beyond the 190-kb *Sal*I fragment. This fragment was 150 kb long, most of it (140 kb) extending beyond the *Sal*I site (Fig. 2). The *Mlu*I–*Bgl*I double digestion reduced the 150-kb *Bgl*I fragment to a 100-kb fragment hybridizing to phaseolin, whereas *Sma*I digestions cut the

same fragment to an even shorter length of 75 kb hybridizing to phaseolin. The *Sal*I phaseolin fragment was located asymmetrically in the large, 450-kb *Sma*I fragment, facilitating the orientation of the locus in the linkage group, as described subsequently.

In plants, storage proteins are usually encoded by multigene families, but the organization and distribution of those genes in the genome is highly diverse. Legumins in pea, for example, are encoded by many genes showing high differentiation into subgroups and cluster in several loci (Turner et al. 1993). Storage proteins in other plants seem to be encoded by clustered gene families with a low degree of differentiation. The patatin storage protein in potato tubers is encoded by 10–15 genes separated by at least 15–20 kb in a single locus of a maximum of 1.4 Mb (Twell and Ooms 1988; Ganal et al. 1991).

The identification of the "I" phaseolin, the presumed ancestral phaseolin sequence in common bean (Kami et al. 1995), provides the starting point for more detailed studies on the evolution of the organization of *Phs* within *P. vulgaris*. A first characteristic of this evolution is the apparent increase in size of the locus from 160 to 190 kb (data not shown). Additional PFGE experiments, possibly combined with tests distinguishing individual phaseolin genes (including the PCR test for tandem repeats used here), should allow us to gain further insights into how the locus has expanded in size.

Physical distances to flanking markers

Restriction digestions with enzymes generating very large (>330 kb) fragments containing phaseolin sequences were used to establish a restriction map to the flanking genetic markers *D1861* and *Bng060* and to orient the locus in the D7 linkage group. These enzymes included *Not*I, *Sfi*I, *Bss*HII, *Eag*I, *Sma*I, and *Sac*II. Probes for the loci *D1861* and *Bng060* are both single-copy sequences of less than

Fig. 2. Long-range restriction map of the phaseolin region. Bg, *Bgl*I; Bs, *Bss*HII; E, *Eag*I; M, *Mlu*I; N, *Not*I; Ns, *Nsp*V; Sc, *Sac*II; Sf, *Sfi*I; Sl, *Sal*I; Sm, *Sma*I; X, *Xho*I. The dark area represents the genomic area occupied by *Phs*; the hatched areas on the left and right of *Phs* are the approximate locations of *Bng060* and *D1861*.
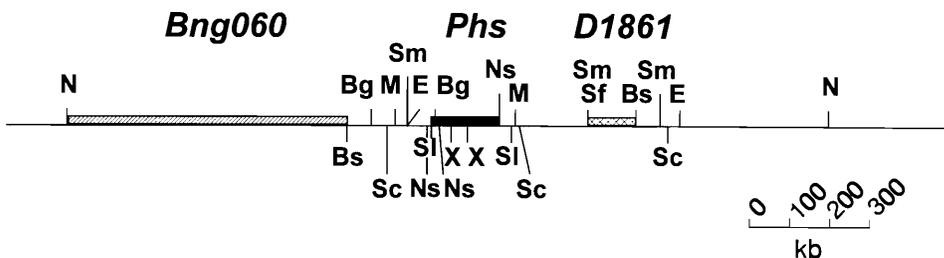
1000 bp. Probes for phaseolin and the *D1861* marker, located 1.9 cM from *Phs* (Fig. 1), hybridized to the same 680-kb *Eag*I and 720-kb *Bss*HII fragments (Fig. 3). In double digestion with *Bss*HII and *Eag*I, the two probes hybridized to a single 560-kb fragment (Fig. 3). The marker *Bng060*, 3 cM from *Phs* in the opposite direction from *D1861* (Fig. 1), hybridized to different *Eag*I and *Bss*HII fragments. However, a double digestion between *Not*I and *Sfi*I gave fragments of identical sizes (1300 kb) for *Bng060* and the fragment hybridizing for phaseolin (Fig. 3). *D1861* hybridized to a different *Not*I–*Sfi*I fragment, approximately 600 kb long. *Bng060* did not hybridize to the same *Sma*I, *Mlu*I, *Sac*II, or *Sal*I phaseolin fragments and must be located at more than 650 kb from *Phs* in the opposite direction of *D1861*. Thus, according to this restriction map, marker *D1861* is located 175–370 kb from the beginning of *Phs*. On the other side of *Phs*, *Bng060* is located at 250–850 kb from the beginning of *Phs* (Fig. 2). The three markers are contained in a single *Not*I fragment of approximately 1900 kb (Fig. 3).

Correlations between physical and genetic distances indicate that recombination in the phaseolin region is comparable to the average for the whole *P. vulgaris* genome. The *P. vulgaris* genetic map covers approximately 1250 cM (Gepts et al. 1993). The *P. vulgaris* haploid genome contains 7×10^8 bp, giving an average relationship between physical and genetic distance of 560 kb/cM. The estimate for the *D1861–Phs* region is 192–290 kb/cM and for the *Phs–Bng060* region is 160–283 kb/cM. This is 1.9–3.0 and 1.6–3.8 times, respectively, higher than the average calculated over the entire genome. If *Phs* is not considered in the calculations, the estimates are higher, about 2.8–6 and 2–6 times higher for the *D1861–Phs* and *Bng060–Phs* intervals, respectively. However, a 3–4-fold variation in recombination frequency falls well within the range observed in *P. vulgaris* (Paredes and Gepts 1995).

Although no work has previously been performed to correlate physical and genetic distances in *P. vulgaris*, research in other plant systems suggests that our results fall within the range of variation for recombination observed for other plants. In tomato, the overall relationship between physical and genetic distance is estimated at 750 kb/cM (somewhat higher than the average for *P. vulgaris*). Nevertheless, the correspondence between genetic and physical distance in tomato can vary from region to region by as much as 100 times. For example, in the *Tm2a* region,

which maps close to the centromere of chromosome 9 in a heterochromatic region, the estimate is 4000 kb/cM (Ganal et al. 1989). Conversely, the tomato RFLP markers *TG105* and *TG26*, 4.1 cM apart distal to the centromere, are located at less than 175 kb from each other (that is, 43 kb/cM or less; Segal et al. 1992). More recently, Wing et al. (1994) estimated the relationship between the physical and genetic distance between *TG523* and *RPD158*, flanking the *jointless* gene in tomato, to be about 200 kb/cM, similar to our estimates for the phaseolin region.

CpG and CpXpG islands, also called HTF (*Hpa*II tiny fragment) islands (Bird 1986; Cheung et al. 1992; Gruenbaum et al. 1981) were evident at both sides of the 190-kb *Sal*I fragment containing all the phaseolin genes. They were detected by a high incidence of restriction sites for endonucleases recognizing unmethylated CpG or CpXpG sequences (Fig. 2). In mammals, it is estimated that 75% of all the *Sac*II, *Bss*HII, and *Eag*I sites and 90% of all the *Not*I sites are located in HTF islands (Lindsay and Bird 1987). In plants, as in mammals, CpG islands are generally associated with the 5′ ends of genes (Gardiner-Garden and Frommer 1992). Consequently, the presence of regions of high incidence of CpGs and CpXpGs in the restriction map of the phaseolin region, including those flanking the *Sal*I fragment, may be indicative of genes expressed in leaves. Although genes controlling interactions between beans and bacteria (*Xanthomonas* and *Rhizobium*) have been mapped to the phaseolin region (Nodari et al. 1993*b*), it is premature to conclude that these genes would be located close to these HTF islands.

Internal map of *Phs*

An internal restriction map of *Phs* was established following digestion with the restriction enzymes *Sal*I, *Mlu*I, *Bst*BI, and *Xho*I (Fig. 2). *Bst*BI was later replaced by its isoschizomer, *Nsp*V, which produced fewer artifactual fragments. In addition, digestion with *Bgl*I, followed by hybridization with a 5′-specific phaseolin probe or with a full length phaseolin probe, was used to establish the orientation of individual phaseolin genes. Data available prior to this study suggested that all phaseolin sequences, not only from *P. vulgaris* but also from *Phaseolus lunatus*, have an internal *Bgl*I site located between the 15-bp (fourth exon) and 27-bp (sixth exon) tandem direct repeats (Kami and Gepts 1994; J. Kami and P. Gepts, unpublished results). This was further verified by PCR analysis of genomic

Fig. 3. Determination of the physical distances from *Phs* to the flanking genetic markers.
(A) Sequential Southern hybridization to phaseolin and Bng060 probes. (B) Sequential
Southern hybridization to phaseolin and D1861 probes. Restriction digestions: lane: 1,
*Eag*I; 2, *Eag*I+*Nsp*V; 3, *Bss*HII; 4, *Bss*HII+*Nsp*V partial digestion; 5, *Sac*II+*Bss*HII;
6, *Sac*II+*Not*I; 7, *Sac*II+*Sfi*I; 8, *Not*I; 9, *Not*I+*Sfi*I; 10, *Sfi*I. The arrows in panel A
indicate the coincident band between phaseolin- and Bng060-probe hybridizations.
Arrows in panel B indicate coincident bands between phaseolin- and D1861-probe
hybridizations. The position of yeast chromosomes, used as molecular weight standards,
is shown on the left.
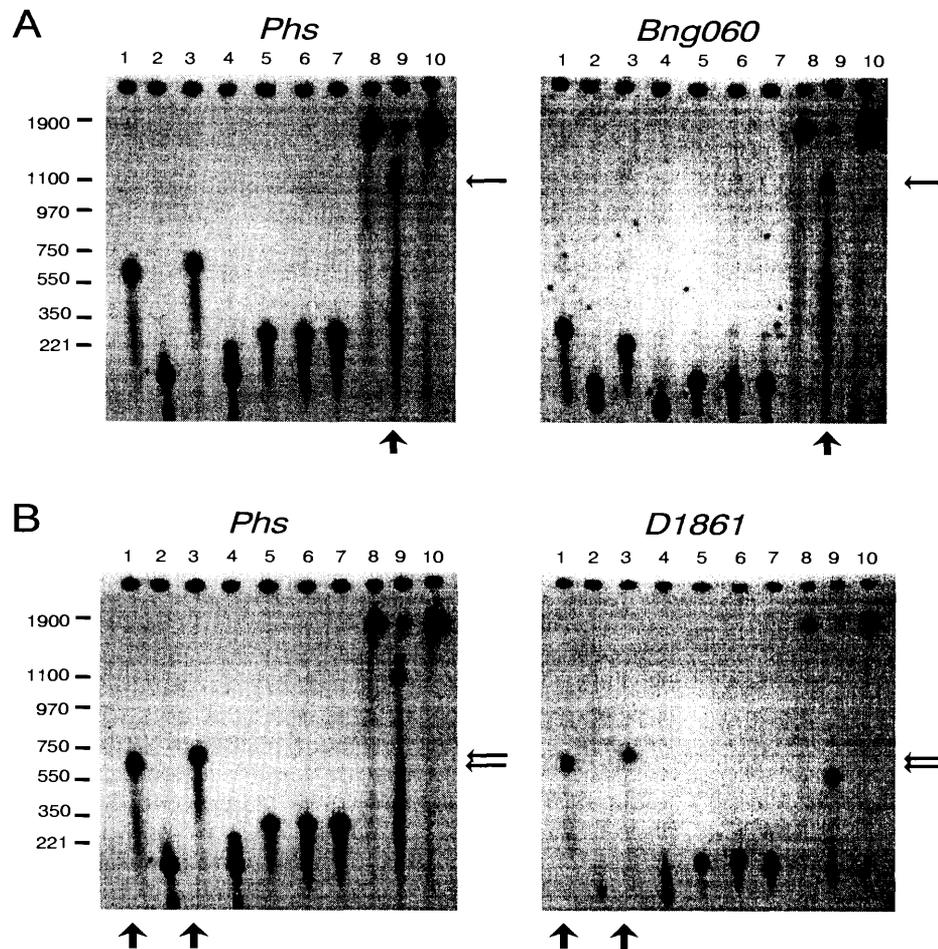
DNA that was either undigested or digested with *Bgl*I.
PCR amplification of uncut DNA from *P. vulgaris* using the
upstream and downstream primers of the 15-bp and the
27-bp diagnostic tests, respectively (Kami et al. 1995),
yielded fragments approximately 900 bp long (not shown).
However, *Bgl*I-digested *P. vulgaris* DNA did not yield any
amplification product using the same set of primers, as
predicted if all the phaseolin-related sequences had a *Bgl*I
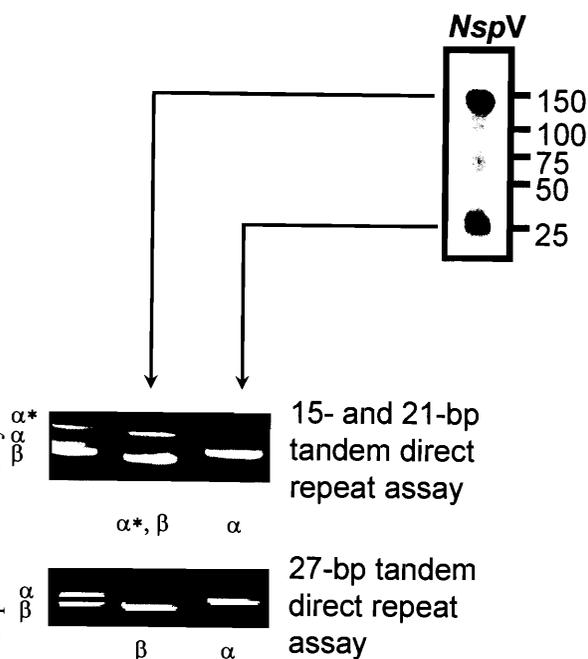site between the sequences recognized by the two primers
(not shown).

The *Bgl*I-digestion pattern in *P. vulgaris* cv. Tendergreen
consisted of four intense fragments and three fragments
with a reduced intensity after hybridization with a 1500-bp
(nearly full length) phaseolin cDNA probe. A radiolabeled
900-bp probe generated by PCR amplification of most of
the 3' half of the phaseolin genes (see previous paragraph)
recognized only the four intense fragments, but not the
three faint fragments, suggesting that the intense fragments

comprise most of the phaseolin genes and that the faint
ones are located at the 3' end of these genes (not shown).

*Nsp*V (Fig. 4) and *Bgl*I (not shown) phaseolin restriction
fragments were separated by PFGE and subjected individually to PCR analysis. PCR products for the 15-bp tandem
direct repeat diagnostic test in *P. vulgaris* cv. Tendergreen
were of three different sizes, 250, 265, and 286 bp (Kami
et al. 1995). The smallest and the medium-sized fragments
corresponded to β- (absence of the 15-bp tandem direct
repeat in exon 4) and α- (presence of the 15-bp tandem
direct repeat in exon 4) phaseolin genes, respectively. The
largest amplified fragment corresponds to an α gene with
an additional 21-bp tandem direct repeat in intron 3 (Kami
et al. 1995).

The large (150 kb) *Nsp*V restriction fragment, expected
to contain the majority of the phaseolin-related sequences
(Fig. 2), contained both α-phaseolin (with the third intron
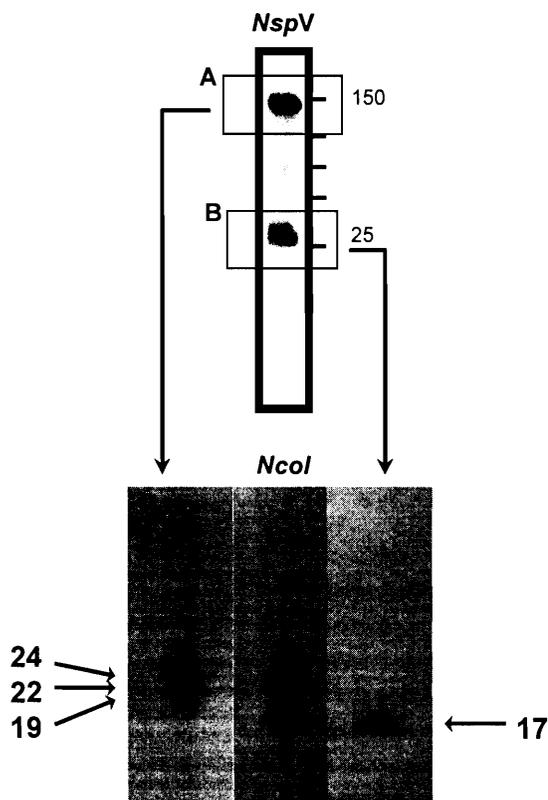21-bp tandem direct repeat) and β-phaseolin genes (Fig. 4).

Fig. 4. PCR analysis of individual NspV restriction fragments recognized by a phaseolin probe. Top panel: Southern hybridization of *P. vulgaris* high molecular weight DNA digested with NspV and hybridized to a phaseolin probe. Lower panels: PCR diagnosis for the 15- and 27-bp tandem direct repeats of the phaseolin genes (Kami et al. 1995) performed in DNA isolated from agarose gels at the positions expected for the two NspV bands shown in the top panel. The patterns in the leftmost lane of the two lower panels are PCR products from *P. vulgaris* cv. Tendergreen genomic DNA as controls.

Fig. 5. Two-dimensional analysis of phaseolin NspV–NcoI double digests. Upper panel: first dimension PFGE separation of NspV-digested high molecular weight DNA. Lower left and right panels: second-dimension PFGE separation following digestion with NcoI; NspV–NcoI double digest fragments are indicated with an arrow. Lower central panel: one-dimensional PFGE separation of high molecular weight DNA following NcoI digestion as control. Numbers indicate molecular weight standards in kilobases.

The small (25 kb) NspV fragment contained only one or more α genes without the 21-bp tandem direct repeat (Fig. 4). The largest *BglI* fragment, mapped by double restriction digestion to the small NspV fragment towards marker *Bng060*, also showed the presence of the 5′ end of a short α gene (i.e., without the third intron 21-bp repeat; not shown). Most of the phaseolin genes in the large NspV fragments did not appear to contain the 27-bp tandem direct repeat (Fig. 4). These observations indicate that the phaseolin gene closest to the *Bng060* marker is an α-phaseolin gene lacking the third intron 21-bp tandem direct repeat (Fig. 2). The orientation of this gene is with the 5′ end towards the *Bng060* marker.

In order to gain additional information about individual genes at *Phs*, we used two-dimensional agarose gel electrophoresis techniques with NspV–NcoI (Fig. 5) and *BglI*–NcoI (not shown) double restriction digestions. Phaseolin genes have no internal NspV site (see above) or NcoI sites, although one NcoI fragment is located 2.5 kb upstream of the transcription initiation point, as deduced from genomic sequences (Hall et al. 1984). Based on the two-dimensional analysis, all phaseolin genes were included in NcoI fragments of four different sizes (Fig. 5). Double digestions using NcoI and *BglI* showed that all the NcoI phaseolin fragments contained one internal *BglI* site located

close to the 3′ end of the coding sequence of the individual phaseolin sequences (not shown). The data obtained from one- and two-dimensional gel electrophoresis and PCR amplifications are consistent with several possible gene arrangements within *Phs*. Resolution of these ambiguities in gene arrangements will require cloning of the entire *Phs* into large-fragment libraries, such as BAC (bacterial artificial chromosome) libraries, followed by restriction mapping of the clones and hybridization to phaseolin probes.

Acknowledgments

(Université de Paris-Sud, Orsay, France) for providing the P clones and information on their high molecular weight DNA extraction procedure, and to Joe Tohme (Centro Internacional de Agricultura Tropical, Cali, Colombia) for supplying Bng clones.

References

Adam-Blondon, A.F., Sevignac, M., Dron, M., and Bannerot, H. 1994. A genetic map of common bean to localize specific resistance genes against anthracnose. Genome, **37**: 915–924.

Anand, R., and Southern, E.M. 1990. Pulsed field electrophoresis. *In* Gel electrophoresis of nucleic acids. *Edited by* D. Rickwood and B.D. Hames. IRL, Oxford. pp. 101–123.

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.

Bird, A.P. 1986. CpG-rich islands and the function of DNA methylation. Nature (London), **321**: 209–213.

Bliss, F.A., and Brown, J.W.S. 1982. Genetic control of phaseolin protein expression in seeds of common bean, *Phaseolus vulgaris*. Qual. Plant. Plant Foods Hum. Nutr. **31**: 269–280.

Brown, J.W.S., Bliss, F.A., and Hall, T.C. 1981. Linkage relationships between genes controlling seed proteins in French beans. Theor. Appl. Genet. **60**: 251–259.

Bustos, M.M., Guiltinan, M.J., Jordano, J., Begum, D., Kalkan, F.A., and Hall, T.C. 1989. Regulation of β-glucuronidase expression in transgenic plants by an A/T-rich, *cis*-acting sequence found upstream of a French bean β-phaseolin gene. Plant Cell, **1**: 839–853.

Casey, R., and Domoney, R. 1987. The structure of plant storage protein genes. Plant Mol. Biol. Rep. **5**: 261–281.

Chee, P.P., Klassy, R.C., and Slightom, J.L. 1986. Expression of a bean storage protein "phaseolin minigene" in foreign plant tissues. Gene (Amst.), **41**: 47–57.

Cheung, W.Y., Moore, G., Money, T.A., and Gale, M.D. 1992. *Hpa*II library indicates 'methylation-free islands' in wheat and barley. Theor. Appl. Genet. **84**: 739–746.

Cramer, J.H., Lea, K., and Slightom, J.L. 1985. Expression of phaseolin cDNA in yeast under the control of natural plant DNA sequences. Proc. Natl. Acad. Sci. U.S.A. **82**: 334–338.

Cramer, J.H., Lea, K., Schaber, M.D., and Kramer, R.A. 1987. Signal peptide specificity in posttranslational processing of the plant protein phaseolin in *Saccaromyces cerevisiae*. Mol. Cell. Biol. **7**: 121–128.

Creusot, F., Sevignac, M., and Dron, M. 1992. Large-scale mapping using high-molecular-weight DNA from *Phaseolus vulgaris*. Plant Mol. Biol. Rep. **10**: 207–218.

Feinberg, A.P., and Vogelstein, B. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**: 6–13.

Ganal, M.W., and Tanksley, S.D. 1989. Analysis of tomato DNA by pulsed field gel electrophoresis. Plant Mol. Biol. Rep. **7**: 17–27.

Ganal, M.W., Young, N.D., and Tanksley, S.D. 1989. Pulsed field gel electrophoresis and physical mapping of large DNA fragments in the *Tm-2a* region of chromosome 9 in tomato. Mol. Gen. Genet. **215**: 395–400.

Ganal, M.W., Bonierbale, M.W., Roeder, M.S., Park, W.D., and Tanksley, S.D. 1991. Genetic and physical mapping of the patatin genes in potato and tomato. Mol. Gen. Genet. **225**: 501–509.

Gardiner-Garden, M., and Frommer, M. 1992. Significant CpG-rich regions in angiosperm genes. J. Mol. Evol. **34**: 231–245.

Gepts, P., Nodari, R., Tsai, R., Koinange, E.M.K., Llaca, V., Gilbertson, R., and Guzmán, P. 1993. Linkage mapping in common bean. Annu. Rep. Bean Improvement Coop. **36**: xxiv–xxxviii.

Gruenbaum, Y., Maveh-Many, T., Cedar, H., and Razin, A. 1981. Sequence specificity of methylation in higher plant DNA. Nature (London), **292**: 260–262.

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.

Kami, J., Becerra Velásquez, B., Debouck, D.G., and Gepts, P. 1995. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. Proc. Natl. Acad. Sci. U.S.A. **92**: 1101–1104.

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

Koenig, R., and Gepts, P. 1989. Segregation and linkage of genes for seed proteins, isozymes, and morphological traits in common bean (*Phaseolus vulgaris*). J. Hered. **80**: 455–459.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M., Lincoln, S.E., and Newburg, L. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics, **1**: 174–181.

Lindsay, S., and Bird, A.P. 1987. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature (London), **327**: 336–338.

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

Mehdy, M.C., and Lamb, C.J. 1987. Chalcone isomerase cDNA cloning and mRNA induction by fungal elicitor, wounding and infection. EMBO J. **6**: 1527–1533.

Murray, M.G., and Kennard, W.C. 1984. Altered chromatin conformation in the higher plant gene phaseolin. Biochemistry, **23**: 4225–4232.

Nodari, R.O., Koinange, E.M.K., Kelly, J.D., and Gepts, P. 1992. Towards an integrated linkage map of common bean. I. Development of genomic DNA probes and levels of restriction fragment length polymorphism. Theor. Appl. Genet. **84**: 186–192.

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

Nodari, R.O., Tsai, S.M., Guzmán, P., Gilbertson, R.L., and Gepts, P. 1993*b*. Towards an integrated linkage map of common bean. 3. Mapping genetic factors controlling host–bacteria interactions. Genetics, **134**: 341–350.

Paredes, O.M., and Gepts, P. 1995. Segregation and recombination in inter-gene pool crosses of *Phaseolus vulgaris* L. J. Hered. **86**: 98–106.

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

Schumann, K., Baumann, A., and Nägl, W. 1990. Localization of phaseolin genes in the polytene chromosomes of *Phaseolus coccineus* (Leguminosae). Genetica, **83**: 73–76.

Segal, G., Sarfatti, M., Schaffer, M.A., Ori, N., Zamir, D., and Fluhr, R. 1992. Correlation of genetic and physical structure in the region surrounding the *Is2 Fusarium oxysporum* resistance locus in tomato. Mol. Gen. Genet. **231**: 179–185.

Sengupta-Gopalan, C., Reichert, N.A., Barker, R.F., and Hall, T.C. 1985. Developmentally regulated expression of

the bean β-phaseolin gene in tobacco seed. Proc. Natl. Acad. Sci. U.S.A. **82**: 3320–3324.

Siedler, H., and Graner, A. 1991. Construction of physical maps of the *Hor1* locus of two barley cultivars by pulsed field gel electrophoresis. Mol. Gen. Genet. **226**: 177–181.

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

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.

Turner, L., Hellens, R.P., Le, D., and Ellis, T.H.N. 1993. Genetic aspects of the organization of legumin genes in pea. Plant Mol. Biol. **22**: 101–112.

Twell, D., and Ooms, G. 1988. Structural diversity of the patatin gene family of potato cv. Désirée. Mol. Gen. Genet. **212**: 325–336.

Vallejos, E.C., Sakiyama, N.S., and Chase, C.D. 1992. A molecular marker-based linkage map of *Phaseolus vulgaris* L. Genetics, **131**: 733–740.

van Daelen, R., and Zabel, P. 1992. Pulsed-field electrophoresis. *In* Plant molecular biology manual. *Edited by* S.B. Gelvin, R.A. Schilperoort, and D.P.S. Verma. Kluwer, Dordrecht, The Netherlands. pp. **C5**: 1–30.

van Santen, V.L., and Spritz, R.A. 1987. Splicing of pre-mRNA in animal systems and vice-versa. Gene (Amst.), **56**: 253–265.

Wing, R.A., Zhang, H.B., and Tanksley, S.D. 1994. Map-based cloning in crop plants—tomato as a model system. 1. Genetic and physical mapping of *jointless*. Mol. Gen. Genet. **242**: 681–688.