Theor Appl Genet (2003) 107:814–822
DOI 10.1007/s00122-003-1343-8

L. Lioi · F. Sparvoli · I. Galasso · C. Lanave · R. Bollini

Lectin-related resistance factors against bruchids evolved through a number of duplication events

Received: 14 January 2003 / Accepted: 7 May 2003 / Published online: 18 June 2003
© Springer-Verlag 2003

Abstract Abundant lectin-related proteins found in common beans (*Phaseolus vulgaris* L.) have been shown to confer resistance against the larvae of a number of bruchid species. Genes encoding for these proteins are members of the lectin multigene family, the most representative components being arcelins, phytohemagglutinins and α-amylase inhibitors. Arcelins have been described in seven variants, some of which are resistance factors against the Mexican bean weevil (*Zabrotes subfasciatus*), a major bean predator. In this study the isolation and sequencing of arcelin genes from wild *P. vulgaris* genotypes, containing Arc3 and Arc7 variants, is reported, and similarities and evolutionary relationships among the seven known arcelins are described. The evolutionary analysis shows that arcelins 3 and 4 cluster together and are the most-ancient variants. A duplication event gave rise to two additional clusters, one comprising arcelins 1, 2 and 6 and separated from the cluster of arcelins 5 and 7. A multiple number of arcelin genes were found in arcelin 3 and 4 genotypes indicating that more than one type of arcelin gene may be present in the same locus. Some of these sequences are reminiscent of ancient duplication events in arcelin evolution demonstrating that arcelins have evolved through multiple duplications. A further aim of this paper was to better understand and describe the evolution of the entire lectin multigene family. Beside arcelins, a number of other types of sequences, such as putative lectins and sequences not easily classifiable, were found in genotypes containing Arc3 and Arc4. These results, together with the evolutionary analysis, indicate that lectin loci are quite complex and confirm their origin by multiple duplication events.

Keywords α-Amylase inhibitor · Arcelin · Evolution · Phytohemagglutinin · Seed storage proteins

Communicated by J.S. Heslop-Harrison

L. Lioi and F. Sparvoli contributed equally to the work

L. Lioi (✉) · I. Galasso
Istituto di Genetica Vegetale, CNR, Via Amendola 165/A,
70126 Bari, Italy
e-mail: lucia.lioi@igv.cnr.it
Tel.: +39-080-5583400
Fax: +39-080-5587566

F. Sparvoli · R. Bollini
Istituto di Biologia e Biotecnologia Agraria, CNR,
Via Bassini 15, 20133 Milano, Italy

C. Lanave
Istituto di Tecnologie Biomediche,
Sezione di Bioinformatica e Genomica, CNR,
Via Amendola 168/5, 70126 Bari, Italy

Introduction

Lectins are sugar-binding proteins that are normally expressed at low levels in many vegetative tissues, but accumulate in a relevant amount in cotyledons, root tubers and bark (Van Damme et al. 1998; Sharon and Lis 1990). In legumes, lectins accumulating in seed tissues are orthologous proteins that are involved in plant defence against bruchids (Chrispeels and Raikhel 1991; Peumans and Van Damme 1995). In the *Phaseolus* genus, the lectin locus has extensively evolved and contains a multigene family coding for up to three major components (Sparvoli et al. 2001). These proteins accumulate in the seed as storage proteins and are involved in the resistance against *Zabrotes subfasciatus* and some *Callosobruchus* species (Huesing et al. 1991; Ishimoto et al. 1996; Osborn et al. 1988; Paes et al. 2000).

In common bean (*Phaseolus vulgaris* L.) the most representative members of the lectin family are phytohemagglutinin (PHA) and α-amylase inhibitor (αAI), that normally occur in wild and cultivated accessions. Arcelin (Arc), the third protein, has been found only in wild accessions from Mexico. Phytohemagglutinin, being the only member of the family that binds carbohydrates, is the true lectin. Arcelin may retain a weak carbohydrate binding-activity whereas αAI is completely devoid of such activity; therefore both will be referred to as lectin-related proteins.

Based on the electrophoretic pattern of its subunits, arcelin has been described in seven allelic variants, composed by a small number of polypeptides in the size range of about 31 to 40 kDa (Osborn et al. 1986; Lioi and Bollini 1989; Santino et al. 1991; Acosta-Gallegos et al. 1998).

Arcelin genes are highly similar and tightly linked to those encoding for PHA and αAI, so they are considered related members of a lectin gene-family (Osborn et al. 1986; Nodari et al. 1993; Mirkov et al. 1994). Analysis of a sequences alignment showed that, compared to PHA, arcelin and α AI present the deletion of one or three short segments, respectively. These deletions result in the lack of one or two loops in the three-dimensional structures of arcelin and αAI (Bompard-Gilles et al. 1996; Mourey et al. 1998). The presence of genes at the same locus, together with the high levels of sequence similarity and gaps in conserved positions, supported the hypothesis that a single ancestor gene gave rise to all the others by tandem duplications and divergence (Rougé et al. 1993; Finardi-Filho et al. 1996).

Recently, we demonstrated that in common bean a lectin ancestor gene underwent a paralogous duplication event, which gave rise to the progenitor of the true lectin and to the progenitor of the lectin-related genes (Sparvoli et al. 2001). The latter evolved originating the active form of αAI, and in some wild genotypes underwent a second duplication event, giving rise to arcelin genes. Therefore, arcelin and αAI originated from a common ancestor and arcelin does not represent a precursor of αAI, as previously supposed on the basis of the presence of gaps in conserved positions. In addition, we suggested that a better understanding of the evolution of the lectin multigene family could be reached by comparing the complexity found in different loci, i.e. members of the lectin family isolated from single genotypes.

To fill the gap in the knowledge of arcelin diversification, here we report on the isolation and sequencing of genes coding for Arc3 and Arc7. These arcelin sequences, together with those of the other arcelins already present in the EMBL databank, allowed us to establish evolutionary relationships within the seven variants.

Among wild genotypes containing arcelin, apparently only those that accumulate Arc3 and Arc4, beside PHA also contain αAI (Santino et al. 1991). Arc4 and αAI genes were isolated from an arcelin 4 genotype (Mirkov et al. 1994), whereas sequences coding for PHA polypeptides were not reported. Here sequences coding for other members of the lectin family from arcelin 3 and 4 genotypes were isolated and used to draw a more-conclusive model on the evolution of the lectin multigene family in *P. vulgaris*. In fact, these genotypes have the most complex lectin locus that, in our opinion, is the most representative for studies on its evolution, since it contains all the three major lectin-related components.

Materials and methods

Plant materials

Wild bean accessions containing arcelin were kindly provided by the CIAT (Centro Internacional de Agricultura Tropical), Cali, Colombia. Accession-identification codes are as follows: G12882 (arcelin 1), G12866 (arcelin 2), G12891 (arcelin 3), G12949 (arcelin 4), G02771 (arcelin 5), G11051 (arcelin 6) and G24591 (arcelin 7).

Protein extraction, SDS-PAGE and Western blotting

Total seed proteins were extracted from mature seeds and separated on 15% SDS-PAGE as described in Bollini and Chrispeels (1978). Gels were stained with Coomassie Brilliant BlueR-250 or blotted on a supported nitrocellulose membrane (BIO-RAD) and Western blots developed according to Burnette (1981), using the rabbit immune serum against Arc3 at 1:1000 dilution and peroxidase-linked anti-rabbit IgG as second antibodies. The immune serum was pre-incubated overnight with 50 μl of a total seed extract from the arcelin-null cv Tendergreen, in 1 ml of phosphate-buffered saline (0.15 M NaCl, 10 mM K-phosphate, pH 7.4), to eliminate any unspecific cross-reactivity (Sparvoli and Bollini 1998).

Southern-blot analysis

Total genomic DNA from the different *P. vulgaris* accessions was isolated from leaves of 10-day old seedlings following the procedure described in Vitale et al. (1985). About 3 μg of genomic DNA were digested with the restriction enzyme *Eco*RI. Agarose electrophoresis and blotting were as in Sparvoli et al. (1996). Clones Pv7-38 (*Arc7*), Pv3-1 (*Arc3-II*) (this paper) and pSC2 (*PHA-L*) (Staswick and Chrispeels 1984), were labelled with [α-^{32}P]-dCTP, and used for filter hybridisation as in Tonelli et al. (1991). Filter washing was in 1.5 mM of sodium citrate pH 7.0, 15 mM of NaCl and 0.5% SDS at 65 °C.

PCR analysis and cloning

Genomic DNA from *P. vulgaris* genotypes G12891, G12949 and G24591 was extracted from 0.1 g of young leaves by the Nucleon protocol (Amersham). The primers P2 and P1 were as described by Mirkov et al. (1994). The resulting PCR products were gel-purified and cloned in pGEM-T (Promega) plasmid. Clones were screened according to the insert size, and to the presence/absence of restriction sites for *Ava*I, *Eco*RV, *Pst*I and *Rsa*I (Fig. 1). Selected clones were sequenced in both directions with an automated sequencer.

Two internal primers specific for the amplification of the *Arc4* gene (Mirkov et al. 1994) were designed as follow: Arc4-for 5′ CCCGACGACCCCGAAGCCCATATT 3′ and Arc4-rev 5′ CTCCATATGCACGCTGCTGGTAGAAGA 3′. PCR conditions were as in Mirkov et al. (1994).

To establish whether the isolated genomic clones were transcribed, cDNA was prepared from polysomal mRNA extracted from mid-maturation cotyledons (Bollini et al. 1983). First-strand cDNA synthesis was done using the "First-strand cDNA synthesis Kit" (Pharmacia), using a *Not*I-d(T)18 primer. RT-PCR was performed as indicated above and PCR products were cloned into pBSKII+ (Stratagene). Clones to be sequenced were selected as above.

Sequence analysis and molecular-evolution methods

Nucleotide sequences were multi-aligned according to their corresponding amino-acid sequence alignments by using the ClustalW program and hand optimised by the SEAVIEW graphic tool

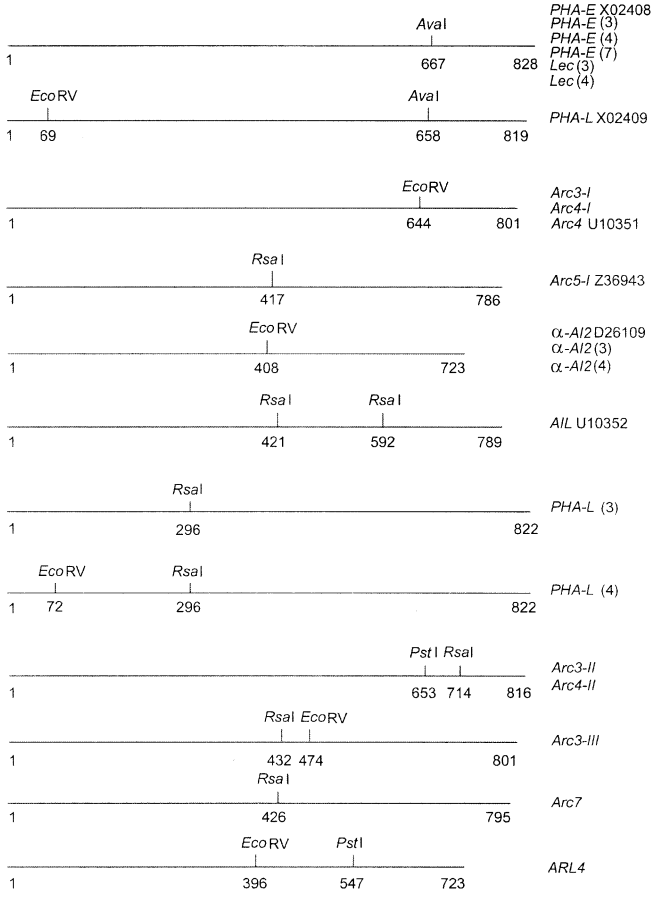

Fig. 1 Restriction maps of lectin and lectin-related genes from *P. vulgaris*. Code names are as in Table 1, except for sequences already present in the EMBL Nucleotide Sequence Database. *Numbers* in parentheses refer to the arcelin genotype

(Galtier et al. 1996). Evolutionary analysis was carried out using the programs implemented in the PAUP* package (Swofford 1998). Genetic distances were calculated by using a stochastic model of gene evolution, the General Time Reversible (GTR) model, also known as the Stationary Markov model (Lanave et al. 1984; Saccone et al. 1990) at non-synonymous positions (first plus second codon position). For the sequences under examination, only the aligned sites with no insertions or deletions were considered. Dendrograms were reconstructed by ML/GTR/NJ-UPGMA methods and the Tree View program was used to draw unrooted and rooted trees. Maximum-likelihood (ML) on evolutionary distances computed on amino-acid sequences by the JTTF method and the neighbor-joining (NJ) reconstruction tree, have been constructed by using PROTML on the MOLPHY package (Adachi and Hasewaga 1994).

Results

Isolation of genes coding for arcelin 3 and 7

In previous work it was proposed that arcelin genes can be clustered into three groups: one comprising *Arc1, Arc2* and *Arc6*, and the other two corresponding to *Arc4* and *Arc5* genes, respectively (Sparvoli and Bollini 1998). To determine the evolutionary relationships among these genes, we cloned the remaining ones, coding for *Arc3* and *Arc7*, and use all the arcelin sequences to make a molecular evolutionary analysis.

Genomic DNA from arcelin 3 and 7 genotypes was used as a template for PCR amplification of the arcelin genes. Since, as expected, the PCR products were a mixture of lectin and lectin-related genes, the strategy chosen to isolate arcelin genes was based either on sequence lengths and analysis of the restriction patterns (Fig. 1).

Several putative arcelin clones were isolated and sequenced. One arcelin sequence was identified in the arcelin 7 genotype (*Arc7*), while two arcelin genes were found in the arcelin 3 genotype (*Arc3-II* and *Arc3-III*). The *Arc7* gene was similar to the *Arc5-III* gene, which displayed an identity of 85.8%. Sequences of the two arcelin 3 clones, *Arc3-II* and *Arc3-III*, showed 86.8% and 84.6% nucleotide identity to *Arc4*, respectively. This was unexpected since the lectin loci in arcelin 3 and arcelin 4 genotypes have a high level of similarity (Osborne et al. 1986; Hartweck et al. 1991), much higher than the similarity among arcelin 1, 2 and 6, which share a gene identity above 98.8% (Sparvoli and Bollini 1998).

To confirm the differences between the two arcelin 3 genes and the published *Arc4* gene, we performed a PCR amplification also in the arcelin 4 genotype. This amplification yielded only one arcelin sequence, named *Arc4-II*, identical to the *Arc3-II* clone. Surprisingly, in arcelin 3 and arcelin 4 genotypes, the *Arc4* sequence could be amplified (*Arc3-I* and *Arc4-I* genes) only when the specific internal primers Arc4-for and Arc4-rev (see Materials and methods) were used coupled to the P1 (3′ end) or P2 (5′ end) terminal primers, respectively. In addition, *Arc4-I* differed from the *Arc4* gene mainly for two frameshifts (insertion at nucleotide 593 and deletion at nucleotide 617) that caused an eight amino-acid change (Fig. 2, boxed amino acids).

All the isolated genes represent expressed sequences, since the same clones were also isolated when the PCR amplifications were performed on cDNAs from maturing cotyledons of arcelin 3, 4 and 7 genotypes.

The derived amino-acid sequences from *Arc3-I, Arc4-I, Arc3-II, Arc4-II, Arc3-III* and *Arc7* showed the deletion of a stretch of eight amino-acids, referred to as GAP_3 (Fig. 2), that is typical of arcelin polypeptides (Mirkov et al. 1994). Generally, arcelins present two or three putative N-glycosylation sites N(X)T/S, where X indicates any amino-acid except proline and threonine which may be substituted by serine. *Arc7* confirms this feature, showing three such sites (Fig. 2, shaded box), whereas both *Arc3-II* and *Arc3-III* present six sites, most of them within the first half portion of the mature polypeptide. Although some of the latter sites are probably not used, due to steric impediments, their number is consistent with the high-molecular-weight shown by arcelin 3 and 4 polypeptide subunits (see below).

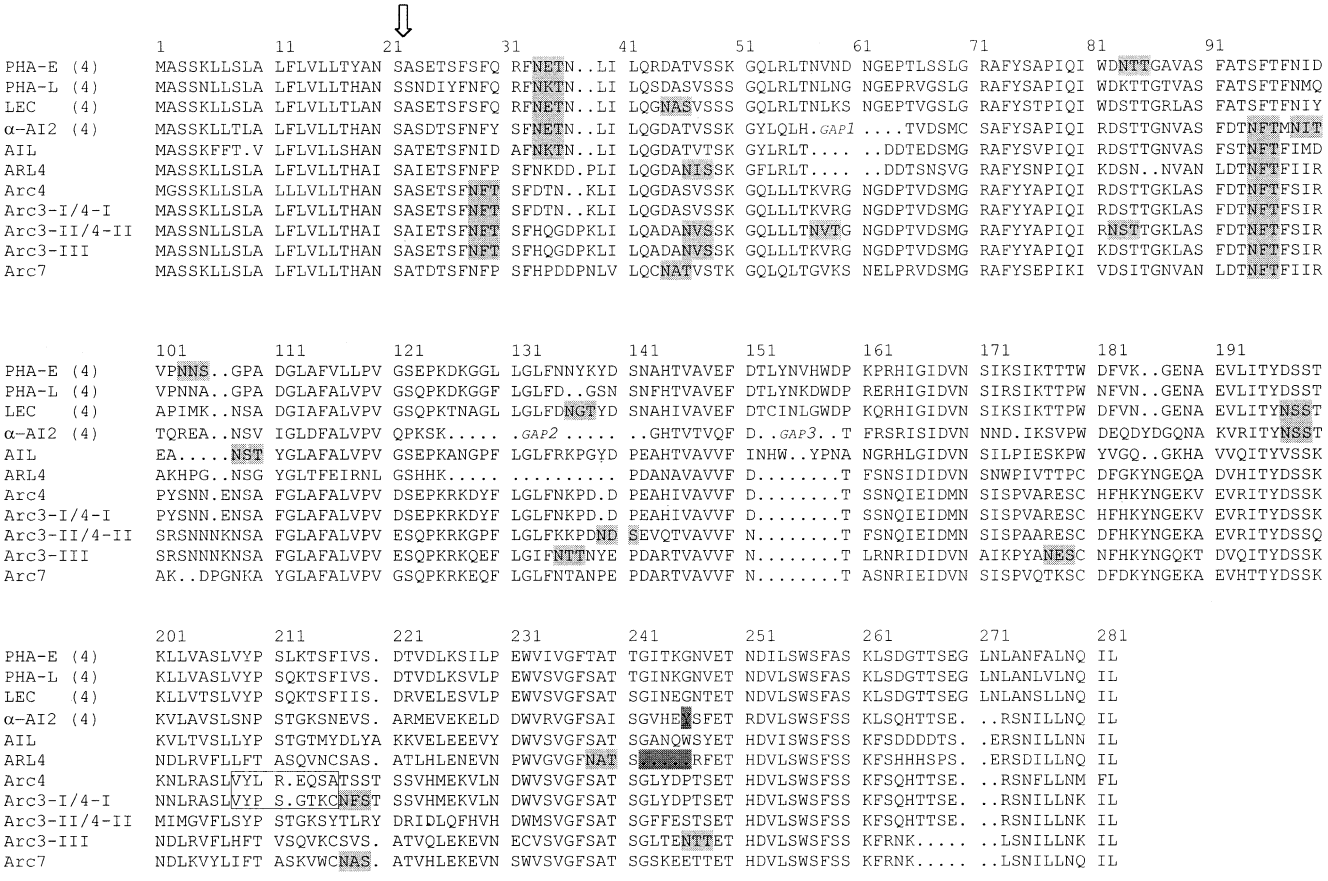

Fig. 2 Multiple alignment of the amino-acid sequences deduced from representative cDNA clones of lectin and lectin-related genes from *P. vulgaris*. Code names of sequences as in Table 1. *Dots* were introduced to maximise similarity. Potential N-glycosylation sites (NXT/S) are in *shaded boxes*. *Dark grey shaded boxes* indicate Y^{186} in mature α AI-2, and the GAP in ARL (see text for explanation). Amino-acid substitutions between Arc4 and Arc4-I are *boxed*. The *arrow* indicates the processing site for removal of the signal peptide. *Numbers in parentheses* refer to the arcelin genotype 4

Arcelin variation assessed throughout SDS-PAGE, Western and Southern blotting

Total protein samples from seeds containing the seven electrophoretic variants of arcelin were submitted to SDS-PAGE under reducing conditions, and immunochemically analysed with polyclonal antibodies raised against Arc3 (Fig. 3A and B). Cross-reaction of the antibodies with lectin-related polypeptides (PHA and αAI) was greatly reduced by pre-incubating the serum with a seed extract containing the latter proteins to block antibodies against common epitopes (Fig. 3B, lane 1) (Sparvoli and Bollini 1998).

The abundance of arcelin polypeptides is comparable in all accessions (Fig. 3A and B) and in some cases greater than that of phaseolin, which is generally the major storage protein in *Phaseolus* seeds (Fig. 3A, compare lanes 5–8 with lanes 2–4). Since arcelin genes code for amino-acid sequences of about 29 kDa, the high-molecular-weight shown by Arc3 and Arc4 polypeptides suggests a higher extent of glycosylation compared to the other arcelins.

Genomic DNA from the seven arcelin variants was digested with *Eco*RI and subjected to Southern-blot analysis using labelled *Arc7, Arc3-II* or *PHA-L* clones as probes (Fig. 4A–C). The hybridisation patterns obtained with the two arcelin probes were almost identical, the major differences being in the hybridisation intensities of specific fragments (Fig. 4A and B). The *Arc7* probe hybridised more strongly than the *Arc3-II* probe to specific fragments in all arcelin genotypes except arcelin 3 and arcelin 4. Indeed, in arcelin 3 and arcelin 4 genotypes the *Arc3-II* probe strongly hybridises to a doublet of about 4.3–4.5 kb which is not specifically recognised by the *Arc7* probe (Fig. 4A, lanes 4 and 5). For both probes one or two more-strongly hybridising fragments were present in each genotype (Fig. 4A and B).

When the *PHA-L* probe was used to re-hybridise the membrane in Fig. 4A a much simpler hybridisation pattern was obtained (Fig. 4C). Two major hybridising fragments were present in all genotypes except arcelin 5, where only one fragment was present, and arcelin 7, which displayed a more-complex pattern. The finding of complex hybridisation patterns when arcelin probes were used, together with the weak hybridisation signals of

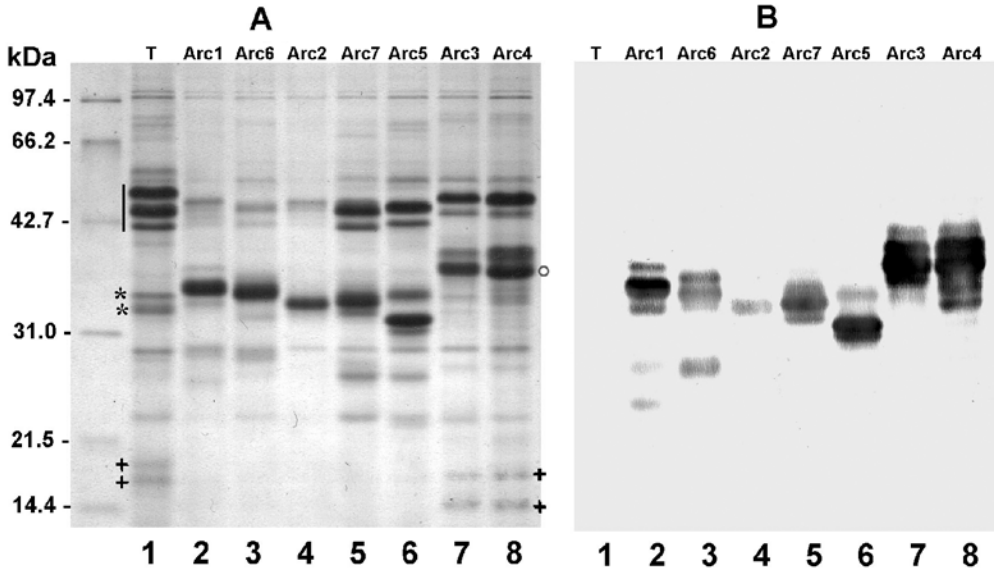

Fig. 3 Comparison of the polypeptide composition of the different arcelin variants. Polypeptides were stained with Coomassie brilliant blue (*panel A*) or subjected to immunoblot analysis with anti-arcelin 3 antiserum (*panel B*). *Vertical bar*: phaseolin polypeptides; *asterisks*: phytohemagglutinin polypeptides; *crosses*: α-amylase inhibitor polypeptides. *T*: Tendergreen cultivar used for comparison. Molecular-weight markers are indicated on the left side

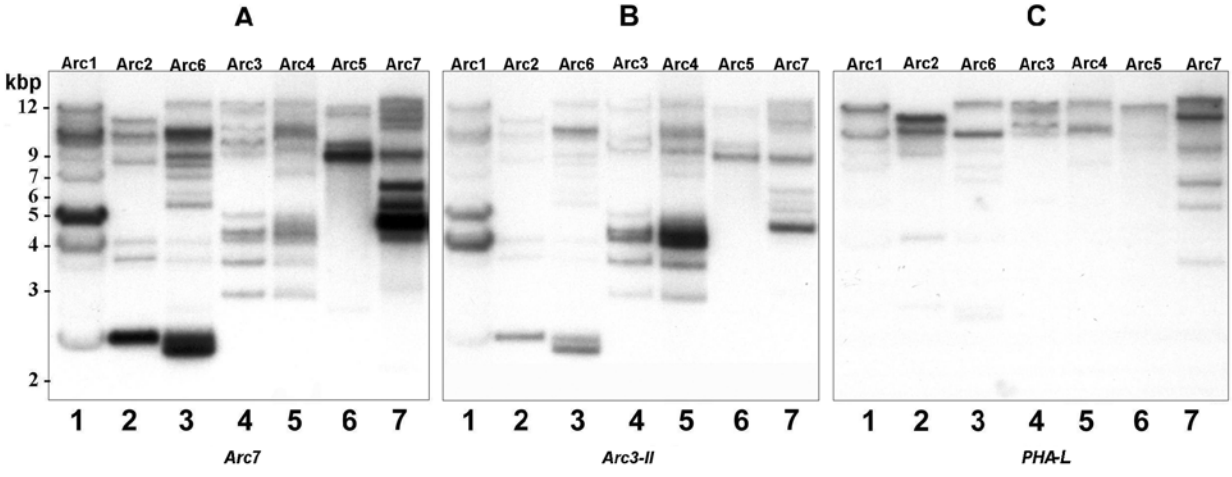

Fig. 4 Southern-blot analysis of the seven arcelin variants. Genomic DNA was digested with the restriction enzyme *Eco*RI and hybridised to the *Arc7* clone (*panel A*), *Arc3-II* clone (*panel B*) or to a *PHA-L* clone (*panel C*)

some fragments, suggest the existence of other lectin-related sequences.

Isolation of other members of the lectin family and sequence comparison

To gain more insights into the complexity of the lectin locus, other members of the lectin family, beside arcelin, were isolated from arcelin 3 and 4 genotypes. These loci express the highest number of lectin-related proteins, since, in addition to arcelin, they also contain αAI and PHA. The analysis was extended also to the arcelin 7 genotype, which lacks αAI, to verify whether the corresponding gene/s is/are silent or at all absent.

Many different clones were obtained and then screened according to their sizes and the presence/absence of cleavage sites for specific restriction enzymes (Fig. 1) This approach resulted in the identification of several other sequences (Table 1). Clones coding for *PHA-E, PHA-L* and *αAI-2* were isolated from arcelin 3 and 4 genotypes. Two additional lectin clones, *Lec3-A1* and *Lec4-B17*, were isolated. These sequences probably code for active lectins, since in their derived amino-acid sequences the amino-acid residues important for metal and sugar binding are conserved (Sharon and Lis 1990; Loris et al. 1998). Sequences amplified from arcelin 3 and arcelin 4 genotypes confirm that their lectin loci are very similar, with a percentage of identity between corresponding genes of about 95.5%. Nevertheless, at least two genotype-specific sequences (*Arc3-III* and *ARL4*) do exist. Based on the conserved features of the lectin-related proteins, the classification of the *ARL4* clone is ambiguous. The presence of the three gaps in conserved positions in the deduced amino-acid sequence suggests it is related to αAIs. Nevertheless, other features contrast

Table 1 Lectin and lectin-related sequences cloned in arcelin 3, 4 and 7 genotypes with respective EMBL accession numbers. NI indicates sequences not identified

Item	Arcelin 3 (G12891)	Arcelin 4 (G12949)	Arcelin 7 (G24591)
Arcelin genes	*Arc3-I* AJ534654	*Arc4-I* AJ439716	*Arc7* AJ439566
	Arc3-II AJ439387	*Arc4-II* AJ532486	NI
	Arc3-III AJ519844	NI	NI
PHA genes	*PHA-E* AJ439562	*PHA-E* AJ439616	*PHA-E* AJ439714
	PHA-L AJ439563	*PHA-L* AJ439617	NI
	Lec AJ439564	*Lec* AJ439715	NI
αAI genes	*αAI-2* AJ439565	*αAI-2* AJ439618	NI
Arcelin-like genes	NI	*ARL4* AJ439619	NI

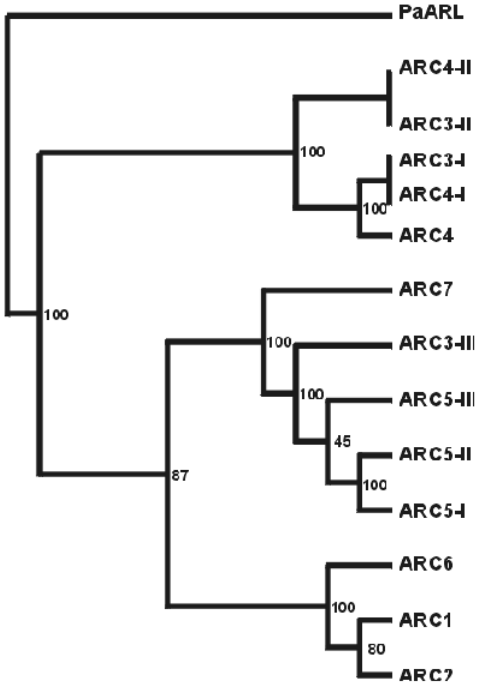

Fig. 5 The phylogenetic tree calculated on the non-synonymous (first plus second) codon positions of nucleotide sequences coding for arcelin variants, by using the General Time Reversible (GTR) model and the UPGMA method. On each node the bootstrap values out of 100 replicates are indicated. The *PaArl* gene was used as an outgroup. EMBL accession numbers of the sequences are reported in Table 1 except for the following: M19430 (*Arc1*); M28470 (*Arc2*); U10351 (*Arc4*); Z36943 (*Arc5-I*); Z36970 (*Arc5-II*); AF193029 (*Arc5-III*); Z36943 (*Arc6*); U10350 (*PaArl*)

with this hypothesis. Firstly, *ARL4* shares a higher amino-acid sequence similarity to arcelins than to αAIs. Secondly, it presents the deletion of a stretch of five amino acids corresponding to Trp^{188} of mature αAI-1. This residue has been suggested to be involved in the functionality of the active site of this inhibitor (Mirkov et al. 1995). In αAI-2, the inhibitor of arcelin 4 genotype, Trp has mutated into Tyr (both the deletion and this residue are shaded in Fig. 2).

No genes coding for PHA-L and αAI were identified in the arcelin 7 genotype, even when genomic DNA was used for amplification.

Evolution of arcelin genes and of the lectin locus in arcelin 3 and 4 genotypes

In arcelin 3 and 4 genotypes three types of arcelin sequences have been identified. Of these, two types have a counterpart in each genotype: the newly identified *Arc3-II* that corresponds to *Arc4-II*, and *Arc3-I* that corresponds to *Arc4-I* and to the previously isolated clone *Arc4* (Mirkov et al. 1994). The third type of arcelin sequence, *Arc3-III*, was instead isolated only in the arcelin 3 genotype.

To address the relationship among the above sequences and the already known arcelins, the gene evolution GTR (general tune reversible) method has been used. The evolutionary distances among the nucleotide sequences encoding for arcelins were used to construct a rooted phylogenetic tree (Fig. 5). Our results show that arcelins can be divided into three main groups. In fact, two duplication events occurred during the evolution of arcelin genes: one separating the precursor of *Arc4, Arc3-I/Arc4-I* and *Arc3-II/Arc4-II* genes from a precursor common to all the other arcelin variants, and the second duplication separating the subgroup of *Arc1, Arc2* and *Arc6* from the subgroup constituted by the *Arc7, Arc3-III* and *Arc5* genes. These findings were confirmed with the unrooted NJ tree on nucleotide sequences and the ML approach on amino-acid sequences (data not shown).

With the aim to draw a more-conclusive model on the evolution of this multigene family, the lectin and lectin-related sequences isolated from the lectin loci of arcelin 3 and 4 genotypes were compared. Phylogenetic trees were constructed combining the above sequences with those already available from the arcelin 4 genotype (*Arc4* and *AIL*). The trees were obtained with ML/GTR/UPGMA-NJ on nucleotide sequences and with ML/JTTF/NJ on amino-acid sequences. In all these trees, two different clusters originating from a duplication event of a lectin progenitor were observed: one grouping the lectin genes and the other one including the lectin-related sequences. In the latter, a second duplication event separated arcelin-related genes from the αAI-related ones (Fig. 6A and B).

Despite the presence of three gaps in conserved positions, the *ARL4* sequence was grouped with the *Arc* genes. In the ML tree shown in Fig. 6B, it is evident that *ARL4* shares a common ancestor to *Arc* genes but most-likely underwent an accelerated process of evolution (see the length of the branch supporting *ARL4*). Similarly, the

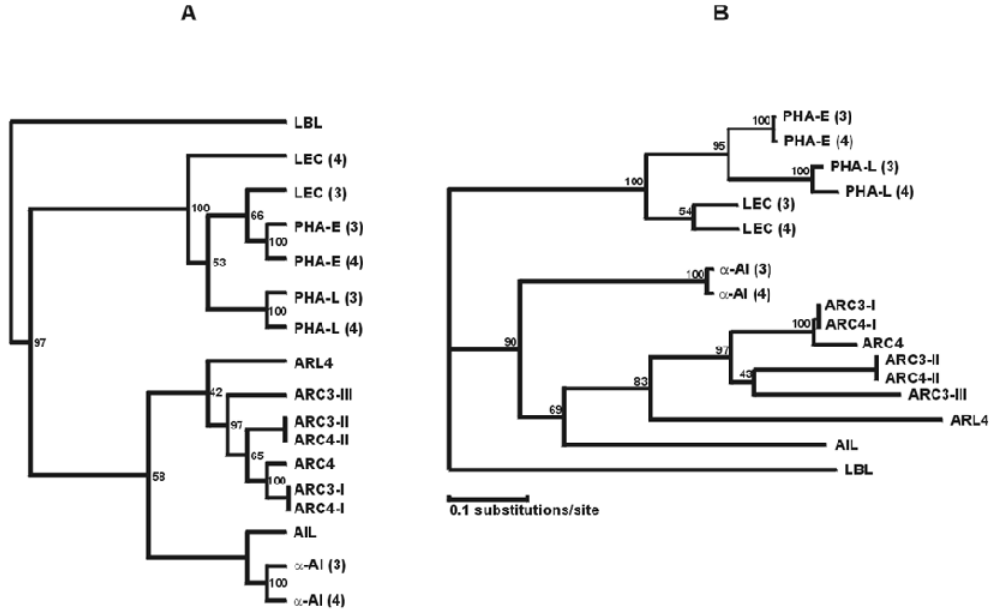

Fig. 6A, B Phylogenetic trees constructed on sequences of members of the lectin family from genotypes containing Arc3 and Arc4. **A** Nucleotide sequences analysed by using the General Time Reversible (GTR) model and the UPGMA method; **B** amino-acid sequences analysed by using the maximum-likelihood method. On each node the bootstrap values out of 100 replicates are indicated. The *LBL* gene was used as an outgroup. The sequence identification reports the abbreviation of the lectin type, and eventually in parentheses, the number of the arcelin genotype. EMBL accession numbers of the sequences are reported in Table 1 except for U10352 (*AIL*), U10351 (*Arc4*) and AJ271873 (*LBL*)

AIL sequence underwent a peculiar evolutionary process leading to a not-well defined position between the two subgroups ARC/αAI (Fig. 6A and B).

Discussion

Common bean has been originally domesticated from wild native plants, but only a narrow segment of the genetic diversity found in nature was captured in the domestication process (Debouck 2000). Many wild *P. vulgaris* genotypes still exist in nature, and could be an inestimable source of desirable traits useful to breeders. Some wild Mesoamerican accessions collected in Mexico are reported to contain arcelin (Acosta-Gallegos et al. 1998). This genetic character can be easily transferred into breeding lines to enhance resistance against the Mexican bean weevil (Cardona et al. 1990).

The main purpose of this study was the isolation and sequencing of arcelin genes from wild *P. vulgaris* genotypes containing Arc3 and Arc7 variants, with the aim to determine similarities and evolutionary relationships among the seven known arcelins.

Generally, arcelin variants are made up by a small number of polypeptides (Osborn et al. 1986; Goossens et al. 1994). In most cases these originate from different post-translational modifications of a single gene product (Young et al. 1999), but may also be the product of different genes. This is the case of the *Arc5* genes (Goossens et al. 1994; Gerhardt et al. 2000). A similar situation is found in the arcelin 1 genotype where at least three types of polypeptides have been reported. Of these, the most abundant is a dimer (Arc1d) and the other two are tetrameric oligomers (Arc1t^1 and Arc1t^2) (Hartweck et al. 1991). It is therefore not surprising that the *Arc4-II* gene codes for a protein having an N-terminus different from that obtained by sequencing the Arc4 protein isolated by Hartweck et al. (1991). Indeed, the latter arcelin is mostly made up by the polypeptide with the highest electrophoretic mobility and represents only a fraction of the total arcelin 4 subunits (in Fig. 3A, lane 8, empty dot). The finding of a multiple number of arcelin genes in arcelin 3 and 4 genotypes further confirms that more than one type of arcelin gene may be present in the same locus.

The analysis of the molecular evolution of arcelin genes resulted in a rooted tree which clearly shows that arcelin genes from arcelin 3 and 4 genotypes cluster together. Furthermore, they were separated from the other arcelin genes by a duplication event that originated two additional subgroups: one comprising *Arc1, Arc2* and *Arc6*, and the other comprising *Arc5, Arc7* and *Arc3-III*. These two subgroups will be referred to as the *Arc1* and *Arc5* subgroup respectively. The *Arc3-III* gene has been identified only in the arcelin 3 genotype. Arcelin 3 and 4 may be considered the most ancient arcelin-containing genotypes (discussed below). The presence of *Arc3-III*, a component of the *Arc5* subgroup, in the arcelin 3 locus, most likely represents a memory of the original duplication event that gave an origin to the precursor of the *Arc1* and *Arc5* subgroups. Reminiscence of the ancient duplication may also be observed in the arcelin 1 locus

(Hartweck et al. 1991). Apparently, this locus expresses, at low-levels, proteins related to Arc3 and Arc4. In fact, the N-terminus of the tetrameric arcelin Arc1t[1] shares a high degree of similarity with the N-terminus deduced from the *Arc3-II, Arc3-III* sequences, and the tetrameric Arc1t[2] shows an even higher similarity to the deduced N-terminal amino-acid sequence of *Arc4* (84.2% of identity) (Hartweck et al. 1991).

Common bean arcelins have been distinguished from the other members of the lectin family on the basis of the presence of GAP3. In fact, lectins have no gaps, whereas αAI, in addition to GAP3, has two more gaps (Rougé et al. 1993; Finardi-Filho et al. 1996; Sparvoli et al. 2001). Our results show that a number of other types of sequences such as putative lectins (*Lec3-A1* and *Lec4-B17*) and sequences not easily classifiable only on the basis of gaps, such as *ARL4*, do exist. At the moment, ARL4, which gene is expressed in maturing cotyledons, still needs to be identified and its biological activity established. Besides *ARL4*, an inactive αAI gene (*AIL*), an amylase inhibitor-like, is believed to be an intermediate form in the evolution of active αAI, that has been reported (Finardi-Filho et al. 1996). Taken together, these findings indicate that an extensive duplication process occurred at the lectin locus, in particular to genotypes containing Arc3 and Arc4.

Our data suggest that genotypes containing either Arc4 or the closely related Arc3 polypeptides, might represent the progenitors of the other arcelin loci, which later during the evolution lost αAI, due to a deletion event. The alternative hypothesis that lack of the αAI protein is due to gene silencing, seems less probable because no αAI clones have been isolated from genotypes containing Arc7. This loss may also have an explanation on a functional basis: the defence role of active αAI might have been replaced by that proposed for arcelins (Acosta-Gallegos et al. 1998). A support for this hypothesis also comes from the observation that lectin and lectin-related polypeptides evolved from tetrameric oligomers (PHA-E, PHA-L, Arc3 and Arc4) towards dimeric proteins (αAI, Arc1, Arc2, Arc5 and Arc7). Moreover, evaluation of genetic distances among different wild *P. vulgaris* genotypes using the AFLP technique, showed that accessions containing Arc3 and Arc4 are in some-way closer to the putative ancestral progenitor than those containing Arc5 and Arc6 (Tohme et al. 1996).

Southern-blot hybridisation confirms the close evolutionary relationships observed in the arcelin 3/arcelin 4 group. Moreover, the multiple hybridisation fragments obtained confirm that arcelin genotypes contain very complex lectin loci.

The further aim of this paper was to better understand and describe the evolution of the entire lectin multigene family: for this reason, representative members of this family were isolated from arcelin 3 and 4 genotypes, known to contain *Arc, PHA* and *αAI* genes (Sparvoli and Bollini 1998). The results confirm that the lectin family has grown through multiple duplication events (Sparvoli et al. 2001). Two different clusters are present, one grouping the PHA genes and the other one grouping the lectin-related sequences. In this latter, a second duplication event separated αAI from the arcelin genes. The absence of arcelin genes from all genotypes, except from some wild Mesoamerican ones, could be explained by assuming that an independent duplication event occurred only in this material. Alternatively, the absence of arcelin in cultivated materials could be explained assuming that arcelin genes appeared in *P. vulgaris*, but changed during evolution giving rise to the active αAI form, and no reminiscent copies remained in *P. vulgaris* accessions, except those found in the Mexican area.

It is interesting to point out that, as for arcelins in the common bean and also in other legume species, the evolution of the lectin locus led to the presence of major lectin-related proteins devoid of sugar-binding activity. This is the case for the closely related ARL (arcelin-like) and AIL, which are by far more abundant that the true Lima-bean lectin (LBL) in the seeds of *Phaseolus lunatus* (Sparvoli et al. 1998), and the most distantly related CLLRP, which accumulates as the second abundant protein in the bark of *Cladrastis lutea* (Van Damme et al. 1995). The biological meaning of the presence of a large amount of lectin-related proteins devoid of sugar-binding activity, and also apparently lacking enzymatic activity, is still puzzling. At present, a role has been described only for the common bean arcelin: the protection of the seed against the Mexican bean weevil is still via an uncleared mechanism (Minney et al. 1990).

Acknowledgements The authors are grateful to Dr. A. Marina Torres, Genetic Resources Unit, CIAT, Cali, Colombia, for providing the *P. vulgaris* accessions used in this study. This work was partially supported by grants of the MiUR Italy project: Cluster C03-legge 488–92.

References

Adachi J, Hasewaga M (1994) MOLPHY programs for molecular phylogenetics, ver. 2.1.2. Institute of Statistical Mathematics, Tokyo

Acosta-Gallegos JA, Quintero C, Vargas J, Toro O, Tohme J, Cardona C (1998) A new variant of arcelin in wild common bean, *Phaseolus vulgaris* L., from southern Mexico. Genet Res Crop Evol 45:235–242

Bollini R, Chrispeels MJ (1978) Characterization and subcellular localization of vicilin and phytohemagglutinin, the two major reserve proteins of *Phaseolus vulgaris* L. Planta 142:291–298

Bollini R, Vitale A, Chrispeels MJ (1983) In vivo and in vitro processing of seed reserve protein in the endoplasmic reticulum: evidence for two glycosylation steps. J Cell Biol 96:999–1007

Bompard-Gilles C, Rousseau P, Rougé P, Payan F (1996) Substrate mimicry in the active center of a mammalian α-amylase: structural analysis of an enzyme-inhibitor complex. Curr Biol 4:1441–1452

Burnette WN (1981) "Western Blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal Biochem 112:195–203

Cardona C, Kornegay J, Posso CE, Morales F, Ramirez H (1990) Comparative value of four arcelin variants in the development

of dry bean lines resistant to the Mexican bean weevil. Entomol Exp Appl 56:197–206
Chrispeels MJ, Raikhel NV (1991) Lectins, lectin genes, and their role in plant defence. Plant Cell 3:1–9
Debouck DG (2000) Biodiversity, ecology and genetic resources of *Phaseolus* beans – seven answered and unanswered questions. Proc 7th MAFF Int Workshop Genet Resour, Part 1 Wild Legumes, AFFRC and NIAR, Japan
Finardi-Filho F, Mirkov E, Chrispeels MJ (1996) A putative precursor protein in the evolution of the bean α-amylase inhibitor. Phytochemistry 43:57–62
Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Comput Appl Biosci 12:543–548
Gerhardt IR, Paes NS, Block Jr C, Mendes PAM, Leite A, Chrispeels MJ, Grossi de Sa MF (2000) Molecular characterization of a new arcelin-5 gene. Biochim Biophys Acta 1490:87–98
Goossens A, Geremia R, Bauw G, Van Montagu M, Angenon G (1994) Isolation and characterisation of arcelin-5 proteins and cDNAs. Eur J Biochem 225:787–795
Hartweck LM, Vogelzang RD, Osborn TC (1991) Characterization and comparison of arcelin seed protein variants from common bean. Plant Physiol 97:204–211
Huesing JE, Shade RE, Chrispeels MJ, Murdock LL (1991) α-Amylase inhibitor, not phytohemagglutinin, explains resistance of common bean seeds to cowpea weevil. Plant Physiol 96:993–996
Ishimoto M, Sato T, Chrispeels MJ, Kitamura K (1996) Bruchid resistance of transgenic azuki bean expressing seed α-amylase inhibitor in common bean. Entomol Exp Appl 79:309–315
Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. J Mol Evol 20:86–93
Lioi L, Bollini R (1989) Identification of a new arcelin variant in wild bean seeds. Annu Rep Bean Improv Coop 32:28
Loris R, Hamerlick T, Bouckaert J, Wyns L (1998) Legume lectin structure. Biochim Biophys Acta 1383:9–36
Minney PHB, Gatehouse MRA, Dobie P, Dendy J, Cardona C, Gatehouse AJ (1990) Biochemical bases of seed resistance to *Zabrotes subfasciatus* (bean weevil) in *Phaseolus vulgaris* (common bean); a mechanism for arcelin toxicity. J Insect Physiol 36:757–767
Mirkov TE, Wahlstrom JM, Hagiwara K, Finardi-Filho F, Kjemtrup S, Chrispeels MJ (1994) Evolutionary relationships among proteins in the phytohemagglutinin-arcelin-α-amylase inhibitor family of the common bean and its relatives. Plant Mol Biol 26:1103–1113
Mirkov TE, Evans SV, Wahlstrom J, Gomez L, Young NM, Chrispeels MJ (1995) Location of the active site of the bean α-amylase inhibitor and involvement of a Trp, Arg, Tyr triad. Glycobiology 3:45–50
Mourey L, Pédelacq JD, Birck C, Fabres C, Rougé P, Samama JP (1998) Crystal structure of the arcelin-1 dimer from *Phaseolus vulgaris* at 1.9 Å resolution. J Biol Chem 273:12914–12922
Nodari RO, Tsai SM, Gilbertson RL, Gepts P (1993) Towards an integrated linkage map of common bean. 2. Development of an RFLP-based linkage map. Theor Appl Genet 85:513–520
Osborn TC, Blake T, Gepts P, Bliss FA (1986) Bean arcelin 2. Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris* L. Theor Appl Genet 71:847–855
Osborn TC, Alexander DC, Sun SSM, Cardona C, Bliss FA (1988) Insecticidal activity and lectin homology of arcelin seed protein. Science 240:207–210
Paes NS, Gerhardt IR, Coutinho MV, Yokoyama M, Santana E, Harris N, Chrispeels MJ, Grossi de Sa MF (2000) The effect of arcelin-1 on the structure of the midgut of bruchid larvae and immunolocalization of the arcelin protein. J Insect Physiol 46:393–402
Peumans WJ, Van Damme EJM (1995) The role of lectins in plant defense. Plant Physiol 109:347–352
Rougé P, Barre A, Causse H, Chatelain C, Porthé G (1993) Arcelin and α-amylase inhibitor from seeds of common bean (*Phaseolus vulgaris*) are truncated lectins. Biochem Syst Ecol 21:695–703
Saccone C, Lanave C, Pesole G, Preparata G (1990) Influence of base composition on quantitative estimates of gene evolution. Methods Enzymol 183:570–583
Santino A, Valsasina B, Lioi L, Vitale A, Bollini R (1991) Bean (*Phaseolus vulgaris* L.) seed lectins: a novel electrophoretic variant of arcelin. Plant Physiol (Life Sci Adv) 10:7–11
Sharon N, Lis H (1990) Legume lectins – a large family of homologous proteins. FASEB J 4:3198–3208
Sparvoli F, Bollini R (1998) Arcelin in wild bean (*Phaseolus vulgaris* L.) seeds: sequence of arcelin 6 shows it is a member of the arcelins 1 and 2 subfamily. Genet Res Crop Evol 45:383–388
Sparvoli F, Daminati MG, Lioi L, Bollini R (1996) In vivo endoproteolytically cleaved phaseolin is stable and accumulates in developing *Phaseolus lunatus* L. seeds. Biochim Biophys Acta 1292:15–22
Sparvoli F, Lanave C, Santucci A, Bollini R, Lioi L (2001) Lectin and lectin related proteins in Lima bean (*Phaseolus lunatus* L.) seeds: biochemical and evolutionary studies. Plant Mol Biol 45:587–597
Staswick P, Chrispeels MJ (1984) Expression of lectin genes during seed development in normal and phytoemagglutinin-deficient cultivars of *Phaseolus vulgaris*. J Mol Appl Genet 2:525–535
Swofford DL (1998) PAUP*. Phylogenetic analysis using parsimony (*and other Methods). Version 4 Sinauer Associates, Sunderland, Massachusetts
Tohme J, Gonzalez DO, Beebe S, Duque MC (1996) AFLP analysis of gene pools of a wild bean core collection. Crop Sci 36:1375–1384
Tonelli C, Consonni SF, Dolfini SL, Dellaporta A, Viotti A, Gavazzi G (1991) Genetic and molecular analysis of *Sn*, a light-inducible, tissue-specific regulatory gene in maize. Mol Gen Genet 225:401–410
Van Damme EJM, Barre A, Bemer V, Rougé P, Van Leuven, Peumans WJ (1995) A lectin and a lectin-related protein are the two most prominent proteins in the bark of yellow wood (Cladrastis lutea). Plant Mol Biol 29:579–598
Van Damme EJM, Peumans WJ, Barre A, Rougé P (1998) Plant lectins: a composite of several families of structurally and evolutionary related proteins with diverse biological roles. Crit Rev Plant Sci 17:575–692
Vitale A, Ceriotti A, Bollini R (1985) Molecular analysis of a phytohemagglutinin-defective cultivar of *Phaseolus vulgaris* L. Planta 166:201–207
Young NM, Thibault P, Watson DC, Chrispeels MJ (1999) Post-traslational processing of two α-amylase inhibitors and an arcelin from the common bean, *Phaseolus vulgaris*. FEBS Lett 446:203–206