

Gabriella Sonnante · Angelo De Paolis  
Domenico Pignone

## Bowman–Birk inhibitors in *Lens*: identification and characterization of two paralogous gene classes in cultivated lentil and wild relatives

Received: 30 July 2004 / Accepted: 21 November 2004 / Published online: 18 January 2005  
© Springer-Verlag 2005

**Abstract** In order to investigate the genetic structure of lentil Bowman–Birk inhibitors (BBIs), primers were designed on pea BBI sequences. The sequences obtained from lentil DNA, using these primers, indicate that lentil possesses at least two paralogous genes. Protein sequences translated in silico from lentil DNA sequences suggest that the two coded proteins are highly similar to *Pisum* trypsin inhibitor TII and TI6 BBIs, respectively. In fact, both are double-headed inhibitors, one class showing the presence of a trypsin- and a chymotrypsin-reactive site, the other showing two trypsin-inhibition sites, similar to pea TII and TI6, respectively. The same primers were used to amplify sequences from the DNA of other *Lens* species. The results strongly support that all species of *Lens* possess the same classes of BBI coding genes, orthologous to those identified in the cultivated lentil. *Lens nigricans* showed the most diverging sequences both at the nucleotide and the amino acid level. The similarity of the two gene classes identified in the genus *Lens* to those of *Pisum* and the observations that the patterns of expression of the *Lens* genes are equivalent to those of pea orthologous genes, possibly imply that BBIs in *Lens* are coded by gene classes with similar genome organization and function to those of pea. Finally, a phyletic analysis, based on the comparison of sequences obtained from other species belonging to the Viciae tribe of the Fabaceae family, strongly suggests that all Viciae could have a similar genome organization and function for BBI genes, and that this could be a general rule in all the Fabaceae family.

Communicated by F.J. Muehlbauer

Publication of the Institute of Plant Genetics N. 50

G. Sonnante (✉) · D. Pignone  
Institute of Plant Genetics, National Research Council (CNR),  
Via Amendola, 165/A, 70126 Bari, Italy  
E-mail: gabriella.sonnante@igv.cnr.it

A. D. Paolis  
Institute of Food Production Science, National Research Council (CNR), Via Prov.le Lecce-Monteroni, 73100 Lecce, Italy

### Introduction

The role of proteinase inhibitors—antinutritional factors affecting animal digestion commonly found in legumes and other plant species—seems to be related to plant defence from attacks by insects, pathogens and other predators (Ryan 1990). There are many families of serine proteinase inhibitors. Among these, the Bowman–Birk type inhibitors (BBIs) are low-molecular-weight molecules with a high cysteine content, these residues being particularly important in maintaining the active conformational structure of the protein so warranting the access to target enzyme (Prakash et al. 1996). BBIs are typically found in several leguminous species, are generally double headed and can therefore inhibit two different proteases per inhibitor molecule (Odani and Ikenaka 1976). Moreover, reports in the last decade show a possible involvement of BBIs in the prevention of tumorigenesis in vitro (Maki et al. 1994; Clemente et al. 2004). These inhibitors are also interesting for the study of host–pathogen co-evolution, as regulators of endogenous enzymes and as markers in studies on plant evolution and diversity (Ryan 1990). Due to their potential utility, the study of variant inhibitors—differing for protein sequence, inhibition performance and selectivity—might be beneficial in identifying new possible sources of profitable molecules.

Pea (*Pisum sativum* L.) is a thoroughly studied legume closely related to lentil, since they both belong to the tribe Viciae. In pea, three classes of genes encoding proteinase inhibitors have been recognised, and genes have been isolated and characterised. The pea trypsin inhibitors (TIs) encoded by these three classes are all members of the same BBI family, but they differ in their sequence, active-site sequence, inhibitor function and patterns of expression; not all of them have the same chromosomal localisation (Domoney et al. 2002). The *TII* gene class is mainly expressed at the level of seed organs, the *TI6* class is expressed in many organs (but is particularly abundant in flowers), and the *TI9* class is

expressed at the level of roots. The pea seed trypsin inhibitors exist in multiple isoforms that can be attributed to post-translational modifications of primary gene products; characterisation of cDNA indicates that two genes are contained in the pea seed TI class (Domoney et al. 1995).

Investigating orthologous genes and proteins can be very useful to understand their functional similarities and differences as well as to achieve information about their evolution. When working with multigene families such as TIs, the problem is to compare orthologous genes between related taxa, paying attention to avoid paralogous comparisons, which are misleading.

Based on recent views, the genus *Lens* Mill. includes seven taxa: the cultivated species (*L. culinaris* Medik. subsp. *culinaris*) including two morphotypes, 'microsperma' and 'macrosperma'; its wild progenitor [*L. culinaris* subsp. *orientalis* (Boiss.) Ponert] and the other wild relatives, *L. odemensis* Ladiz., *L. ervoides* (Brign.) Grande, *L. nigricans* (M. Bieb.) Godr.; and the two recently recognised species *L. tomentosus* Ladiz. and *L. lamottei* Czefr. (van Oss et al. 1997). The cultivated lentil was domesticated in the Old World, but at present its cultivation is spread all over the globe. The six wild *Lens* taxa are mostly distributed in the Mediterranean region, and the lentil wild progenitor reaches Central Asia (Ladizinsky et al. 1984).

The genetic diversity and relationships within the genus *Lens* have been studied using different approaches and markers; these studies have often led to diverging conclusions, possibly due to different evolutionary rates of the genetic systems investigated and/or to the genomic distribution of the single markers utilised in the analyses (Ladizinsky et al. 1984; Pinkas et al. 1985; Hoffman et al. 1986; Havey and Muehlbauer 1989; Muench et al. 1991; Mayer and Soltis 1994; Sharma et al. 1996; Ahmad et al. 1997; van Oss et al. 1997; Ferguson et al. 2000; Sonnante et al. 2003).

In the present study we compare the BBI coding sequences and their presumed coded proteins in the genus *Lens*, in order to (1) ascertain whether there are different classes of TI in lentil; (2) analyse nucleotide and protein sequence differentiation within the genus; (3) compare TI orthologous and paralogous coding DNA and derived protein sequences, also in comparison with BBI sequences from pea and other genera of the Fabaceae; and (4) evaluate the expression of the genes so identified in different organs of cultivated lentil.

## Materials and methods

DNA was extracted from leaves from the plant material indicated in Table 1.

DNA and protein databases were queried to obtain the nucleotide and amino acid sequences of protease inhibitors in some leguminous taxa, and the sequences were aligned using the multialignment program Clustal W,

**Table 1** List of the material analysed

Species	Variety/accession number <sup>a</sup>
<i>Lens culinaris</i> subsp. <i>culinaris</i> macrosperma	cv. Laird
<i>L. culinaris</i> subsp. <i>culinaris</i> macrosperma	Altamura landrace
<i>L. culinaris</i> subsp. <i>culinaris</i> microsperma	cv. Eston
<i>L. culinaris</i> subsp. <i>culinaris</i> microsperma	Castelluccio landrace
<i>L. culinaris</i> subsp. <i>orientalis</i>	PI572366
<i>L. culinaris</i> subsp. <i>orientalis</i>	PI572391
<i>L. culinaris</i> subsp. <i>orientalis</i>	W17080
<i>L. tomentosus</i>	ILWL93
<i>L. tomentosus</i>	ILWL282
<i>L. odemensis</i>	PI572360
<i>L. odemensis</i>	PI572361
<i>L. odemensis</i>	PI572364
<i>L. lamottei</i>	ILWL14
<i>L. lamottei</i>	ILWL29
<i>L. lamottei</i>	ILWL434
<i>L. ervoides</i>	PI572311
<i>L. ervoides</i>	PI572317
<i>L. ervoides</i>	PI572329
<i>L. nigricans</i>	PI572351

<sup>a</sup>Accession numbers are those provided by USDA (*PI*- and *W*- prefixes) or by ICARDA (*ILW*- prefixes)

available online at the European Bioinformatics Institute (EBI) Web site (<http://www.ebi.ac.uk/clustalw/>).

For primer design, nucleotide sequences of EMBL accessions corresponding to pea TI1, TI6 and TI9 (AJ276900, AJ296170, AJ291493, AJ296169, AJ296171), as reported in Domoney et al. (2002), were aligned using the default parameters of the Clustal W program.

Two forward primers (F1 and F2) were designed degenerating nucleotides, which are different in pea *TI1* and *TI6* sequences; a third forward primer (F3) was designed on the sequence of pea *TI1*. The primer F1 (ATGGWGTGAWGAAYAAGAAG) was constructed at the N terminus of the pea TI pre-sequence; F2 (TTGRGCTTCRCCGCAAMT) in the pre-sequence and F3 (GGTGATGATGTCAAATCAG) at the beginning of the mature protein.

Four reverse primers were additionally designed: one within the mature protein (R1: GTGRCATGYTTYATAGCAGAAST); another, R2 (TCAGTTCTTRATKACCTCCTC), at the end of the mature protein of the major seed pea TI1 (accession X83210); a third one (R3: TACATAGTTATACTTGCTCA) just outside the mature protein of the same pea sequence; and the fourth one (R4: GTTTCATCACTTGCATGGT) in the same position as R3, but it was designed on a pea cDNA sequence (accession AJ296169) homologous to the *TI6* gene. Other reverse primers were designed around the end of the mature TI1 pea protein in order to amplify the entire *TI1*-like coding sequence in the cultivated lentil.

Subsequently, the amplification reactions were performed as follows. A total volume of 50 µl for each

reaction contained 50 ng DNA, 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 1.5 mM MgCl<sub>2</sub>, 0.1 mM each dATP, dCTP, dGTP, dTTP, 2.5 U of *Taq* DNA polymerase (Amersham, Uppsala, Sweden) and 0.2 μM primer. The amplification program included a cycle at 94°C for 2 min; 35 cycles at: 94°C for 45 s, 52°C for 45 s, 72°C for 90 s; and a final elongation at 72°C for 5 min. After visualisation on an agarose gel, the PCR product was purified using a Qiaquick PCR purification kit (Qiagen, Hilden, Germany) and directly sent for sequencing. Amplifications and sequencing were repeated three to five times for each primer combination.

Sequences were deposited in the EMBL database (accession numbers AJ420109–AJ420112, AJ420908, AJ420909, and AJ783997–AJ784007).

Nucleotide sequences were translated *in silico* using the program Transeq, available online at the site <http://bioweb.pasteur.fr/seqanal/interfaces/transeq.html>. Sequences were then aligned using the program Clustal W.

Raw nucleotide and amino acid sequences were submitted to cladistic analysis using the MEGA, version 2.1 (Kumar et al. 2001), package, applying the neighbour-joining algorithm and the 'number of differences' model.

Total RNA of *L. culinaris*, cv. Laird was isolated from fresh leaves and germinating seed roots, using Trizol Reagent (Gibco–BRL, Rockville, Md., USA), and from immature seeds using the kit Nucleospin (MACHEREY-NAGEL, Düren, Germany), following the instruction manual. Total RNA was visualised on 1.7% agarose gel. First-strand cDNA was synthesised using SuperScript System for RT-PCR (Invitrogen, Carlsbad, Calif., USA) according to the producer's instructions.

Primer combinations F1-R1 and F1-R4 were used to amplify cDNA from the different organs, using the same PCR conditions as for genomic DNA. Amplified fragments were separated on 1.5% agarose gel.

## Results

### Comparative DNA sequence analysis

The primers were designed starting from pea TI sequences, since the genus *Pisum* is phylogenetically close to the genus *Lens*; in fact they both belong to the tribe Viciae, together with *Vicia* and *Lathyrus*. Amplifications and sequencing were repeated three to five times for each primer combination in order to be sure that the differences observed were not due to sequencing errors. We always obtained very clear and unambiguous sequencing profiles.

In cultivated lentil, each of the three forward primers together with the reverse R1 produced one single band (for each combination) of about 300, 250, and 200 bp, respectively. The first combination of primers (F1-R1) was used to amplify also DNA from the wild *Lens* taxa and produced one band in all

samples, except for *L. nigricans*. For this species, neither the first forward primer (F1) nor the second (F2) produced any amplification product, even when changing the experimental conditions; we also tried a more-degenerated primer designed in the same region as F1, with no successful result. A good amplification in *L. nigricans* was obtained only using the third forward primer (F3) designed at the beginning of the mature protein. The purified amplification products from the first combination of primers were sequenced and produced a sequence of 312 bp.

The primers R2 and R3, expected to allow the amplification of the entire coding region, failed to produce reliable amplifications in combination with any of the forward primers.

The amplification with F1 and R4 produced a single band of 350 nucleotides in all the *Lens* samples.

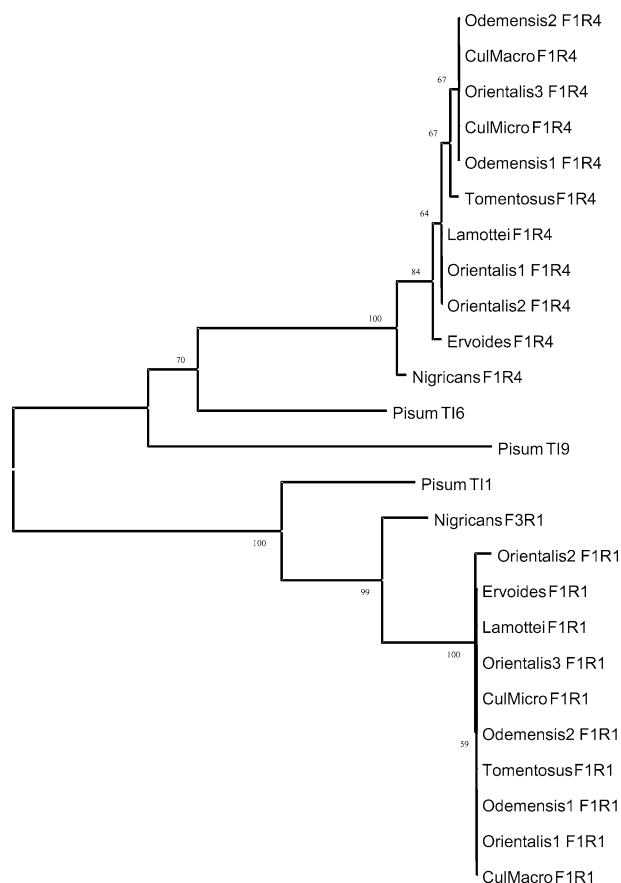
The sequences obtained in cultivated lentil were submitted to FASTA3 for similarity search in the EMBL–EBI Web site. The sequence produced using primers F1 and R1 in cultivated lentil showed the highest similarity (91.3%) with the pea *TI1* gene (AJ276900), whereas it showed 87.2% similarity with the pea *TI9* gene (AJ291493), and 78.5% with the *TI6* gene (AJ296171). Deduced precursor and protein sequences showed the highest similarity values (85.6%) with pea *TI1* protein sequence (Q9M3X3).

On the other hand, the nucleotide sequence obtained with F1 and R4 primers showed the highest similarity (89.4%) with the pea *TI6* gene (AJ296171), whereas the similarity with the *TI9* gene (AJ291493) and with the *TI1* gene (AJ276900) was 80.4% and 76.3%, respectively. Comparing the derived amino acid sequences, we observed the highest similarity (80.8%) with the *TI6*-derived protein sequence (Q8AVS0).

The similarity between the nucleotide sequences obtained in cultivated lentil with F1–R1 primers on one side and with F1–R4 primers on the other was 76.3%; the similarity between the two derived protein sequences was 64.4%.

The nucleotide sequence obtained with F1–R1 primers (F3-R1 in the case of *L. nigricans*) showed few differences among the taxa of the genus *Lens*. As a general rule, the nucleotide sequence was highly conserved in the *Lens* gene pool, with the exception of *L. nigricans*, which showed eight autapomorphic nucleotide substitutions as compared to the cultivated species. The remaining taxa occasionally showed one autapomorphy, sometimes leading to amino acidic substitutions in the deduced protein. The sequence amplified from *L. nigricans* was shorter than those amplified from the other taxa due to the use of a more-internal primer; this implies that it is possible that the number of nucleotide substitutions could be higher in the region not amplified, as also supported by the fact the primer F1 did not give rise to any amplification.

A different pattern arises from the analysis of the DNA sequences amplified using the F1–R4 primer combination (see Fig. 2). In this case the number of



**Fig. 1** Neighbour-joining tree of nucleotide sequences amplified in *Lens* species, using primer combinations F1R1 and F1-R4, compared with three TI sequences from *Pisum*. *Odemensis1 F1R1*: PI572360=PI572361, *Odemensis2 F1R1*: PI572364, *Odemensis1 F1R4*: PI572360=PI572364, *Odemensis2 F1R4*: Od2361, *Orientalis1*: W17080, *Orientalis2*: PI572366, *Orientalis3*: PI572391, *CulMicro*: *L. culinaris* subsp. *culinaris* microsperma, *CulMacro*: *L. culinaris* subsp. *culinaris* macrosperma. Sequences that were identical in the same species were reported only once

nucleotide substitutions appears to be higher and more distributed over the taxa; in these amplicates *L. nigricans* showed only three specific nucleotide substitutions respect to the cultivated species, while sharing two substitutions with other species.

When the sequences were submitted to cladistic analysis based on the mere number of substitutions, without any regard to their type, the nucleotide substitutions present in the sequences generated using the F1–R1 primer combination (F3–R1 for *L. nigricans*) were not distinct enough to generate clusters. As shown in Fig. 1, all *Lens* taxa, except *L. nigricans* (which forms a separate branch), fall in the same cluster. Conversely, those generated with the primer combination F1–R4 resulted in a more-assorted clustering of the species, even though it is clear the lower distance of the most divergent sequence (*L. nigricans*) from the cultivated species as compared to F1–R1 sequences (Fig. 1). It is also evident that these latter sequences form a clade completely distinct from that produced by the sequences amplified using F1–R1 primers.

Moreover, when the two lots of sequences were compared with pea TI sequences, the F1–R1 group clustered with pea *TII* type sequences, while the F1–R4 ones clustered with the *TI6* gene (Fig. 1). These results undoubtedly indicate that the sequences amplified by the primer combinations F1–R1 and F1–R4 do not belong to the same gene; they should therefore be considered paralogous sequences. This view is also supported by the observation that the two genes evolved in the genus *Lens* at different rates, thus indicating lack of homogenisation that should be expected for orthologous genes (Small and Wendel 2002).

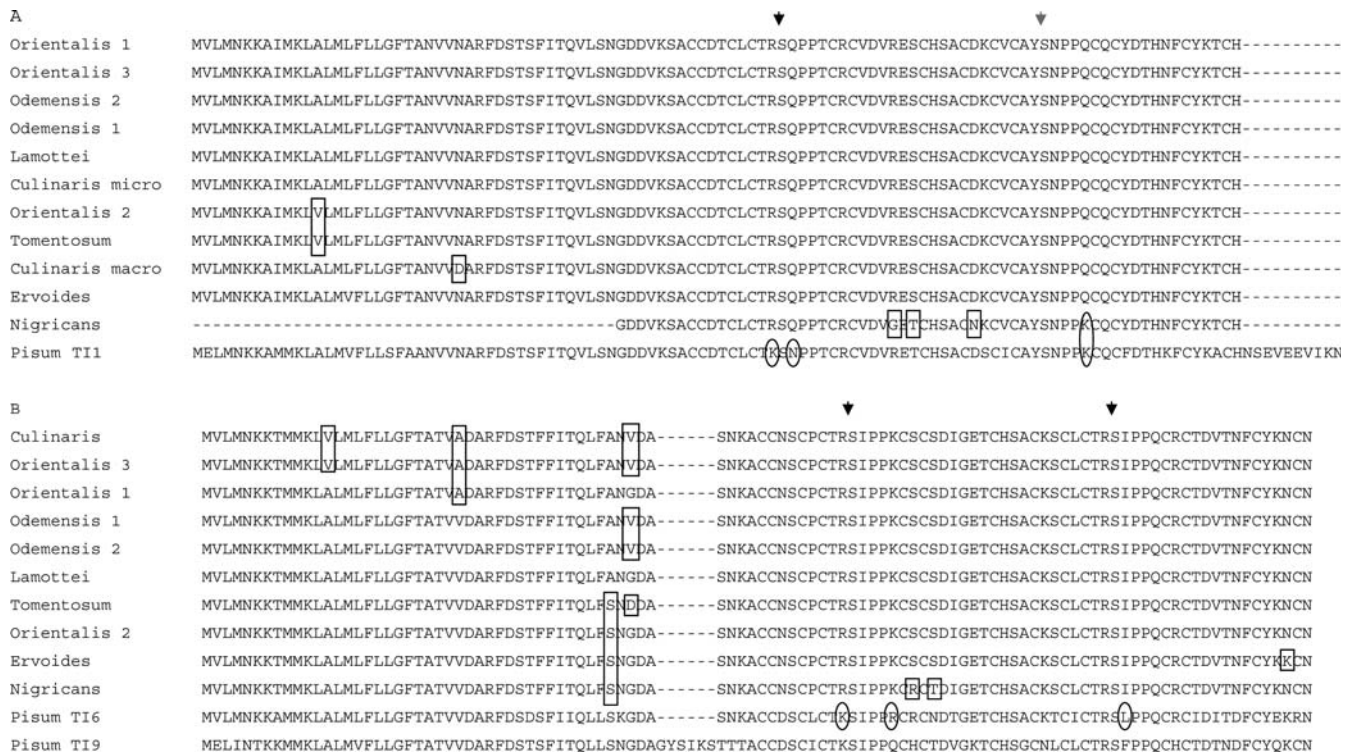
As a general rule, all the sequences from different samples of the same taxon are identical, with some exceptions. Where F1–R1 sequences are concerned, in fact, the *L. culinaris* subsp. *culinaris* samples showed an autapomorphic transition A → G in the samples of the morphotype ‘macrosperma’; in *L. odemensis*, a synonymous nucleotide substitution was observed in one sample only; the three *L. culinaris* subsp. *orientalis* were not homogeneous. As far as the F1–R4 sequences are concerned, a greater homogeneity was observed within the taxa with the sole exception of *L. culinaris* subsp. *orientalis*, which showed intra-population variation.

#### Comparative amino acid sequence analysis

The *in silico* translation of the DNA sequences resulted in rather uniform protein sequences over the different *Lens* species (Fig. 2).

The *Lens* BBIs obtained using the F1–R1 primer combination seem to maintain the same structure as pea major seed TIs (Fig. 2a), which are synthesised as precursor polypeptides, with a precursor of 42 amino acids before the N-terminal glycine, determined for the mature protein (Domoney et al. 1995). This extension is considerably longer than in other inhibitors of the Bowman–Birk class (Rohrmeier and Lehle 1993), even though for a cowpea inhibitor a long precursor has been described (Hilder et al. 1989). In *Lens* pre-sequences, if we exclude *L. nigricans*, only three amino acid substitutions were scored when using the sequences of the cultivated lentil as a reference. The above-mentioned autapomorphic transition A → G observed in the morphotype macrosperma of *L. culinaris* subsp. *culinaris* resulted in an amino acid substitution from asparagine to aspartic acid (see box in Fig. 2a). The sequence of the mature protein deduced by the F1–R1 lot of sequences was identical in all *Lens* taxa, except for *L. nigricans*, which revealed four amino acid substitutions in the mature protein. Also these polypeptides are double-headed inhibitors with a high cysteine content in conserved positions with respect to other known BBIs (Brauer et al. 2001; Mello et al. 2003). The deduced active sites for trypsin and chymotrypsin were identified by comparing the amino acid sequence to the sequence from pea in the literature: the P1–P1' active sites (RS and YS, respectively)





**Fig. 2** Multiple sequence alignment for **a** F1–R1 class and **b** F1–R4 class

are highly conserved in *Lens*, as well as the whole active region included between the cysteine residues in P3 and P6'. Sole exception was a sequence from *L. nigricans*, which showed in the chymotrypsin site a lysine residue instead of a glutamine in position P5'. Comparing *Lens* and pea T11 sequences, in pea the P1 residue of the trypsin active site is a lysine (K) instead of an arginine (R), and the P2' is an asparagine (N) instead of a glutamine (Q), in addition to a difference in position P5' of the chymotrypsin region, which pea shares with *L. nigricans* (circled in Fig. 2a).

When the sequence of the proteins deduced by the F1–R4 lot of sequences was analysed (Fig. 2b), a slightly greater variation was detected, as it could be expected on the basis of the observed higher variation in the nucleotide sequences. Also in this case it was possible to detect the conservation in *Lens* of the two trypsin-inhibitor P1–P1' active sites (RS), as well as of the region included between the cysteine residues in P3 and P6'. This group of amino acid sequences showed—as indicated also by the affinity of nucleotide sequences—a higher similarity to the TI6 protein from pea. Nevertheless there are some amino acid substitutions in the region of the active sites: in the first region the pea TI6 sequence shows differences in P1 and P5' residues, and in the second one a different residue in the P2' position is observed (circled in Fig. 2b).

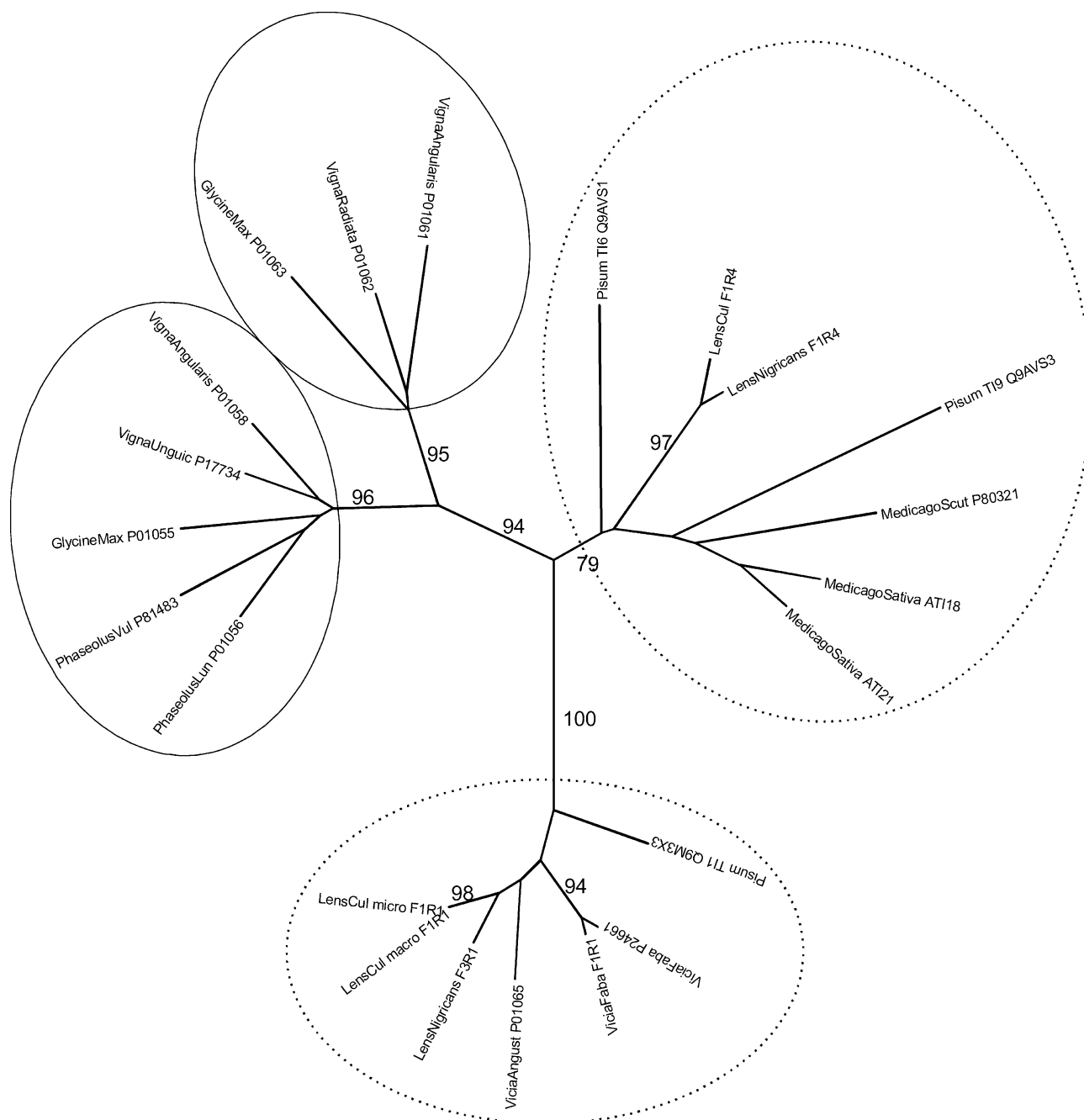
When amino acid sequences were submitted to cladistic analysis, as for the nucleotide sequences, they tended to form two distinct clusters, one showing more

affinity with pea T11 protein and the other with TI6 polypeptide; the structure of the tree obtained from protein sequences is very similar to that obtained from DNA data (data not shown).

#### Analysis of BBI amino acid sequences in some members of the Fabaceae family

The BBI amino acid sequences of *Lens* were compared also with sequences from taxa of the family Fabaceae, namely *Vicia faba* (accession P24661), *V. angustifolia* (accession P01065), *Medicago sativa* (ATI18 and ATI21), *M. scutellata* (accession P80321), *Glycine max* (accessions P01055 and P01063), *Phaseolus vulgaris* (accession P81483), *P. luteus* (accession P01056) *Vigna unguiculata* (accession P17734), *V. angularis* (accessions P01058 and P01061), and *V. radiata* (accession P01062), were retrieved from the databases in order to make a comparison with the sequences from *Lens* and *Pisum*. Although the amino acid sequences may vary considerably, the sequence of the reactive sites for trypsin and chymotrypsin only show some specific variants confined to a limited number of alternative amino acids; moreover the position of the reactive site with respect to C amino acids in P3 and P6', as well as that of the other cysteine residues outside the reactive site, is also strongly conserved over a wide range of genera and families (Mello et al. 2003).

When the available sequences were submitted to cladistic analysis (Fig. 3), bootstrap values clearly evidence that:



**Fig. 3** Neighbour-joining tree of amino acid sequences in some members of the family Fabaceae. Species names are followed by accession numbers as in the EMBL database. Only significant bootstrap values ( $> 50$ ) are reported. *LensCul macro*: *Lens culinaris*

subsp. *culinaris* macrosperma, *LensCul micro*: *Lens culinaris* subsp. *culinaris* microsperma, *ViciaAngust*: *Vicia angustifolia*, *MedicagoScut*: *Medicago scutellata*, *VignaUnguic*: *Vigna unguiculata*, *PhaseolusVul*: *Phaseolus vulgaris*, *PhaseolusLun*: *Phaseolus lunatus*

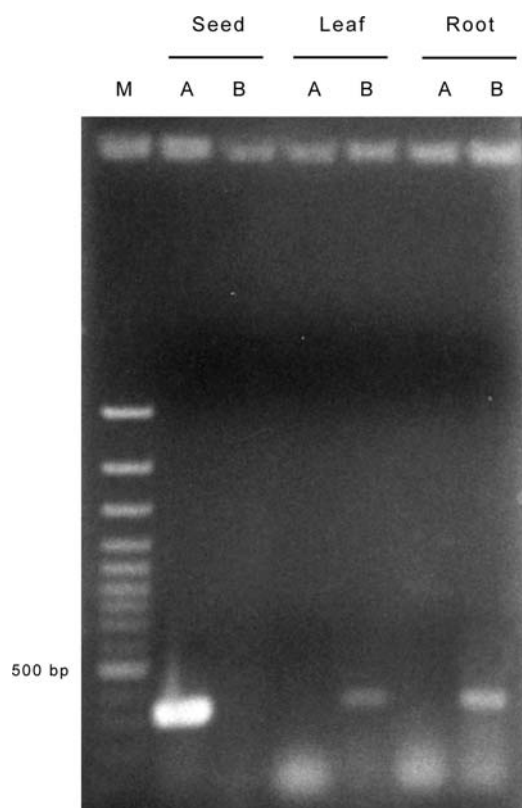
1. Two groups of protein inhibitors are observed within the Viciae tribe, one corresponding to pea TI1 and the other to pea TI6, including pea TI9 which seems to retain a fair degree of similarity to TI6.
2. The BBIs from *Medicago* (tribe Trifolieae) are included in the pea TI6–TI9 group.
3. The BBIs from the genera belonging to the Phaseoleae tribe (*Glycine*, *Vigna*, and *Phaseolus*) are different from the other Fabaceae and form a separate cluster.
4. The Phaseoleae cluster is further subdivided into two more-homogeneous groups, and at least *V. angularis* and *G. max* show to possess two types of proteinase inhibitors that pertain to the different subclusters.

The lentil TI genes and inference on their expression

The entire sequence of the mature TI6-like protein of lentil was obtained, while this was not achieved for the

TII-like lentil protein. In fact, the reverse primer R1 had been designed within the mature protein; therefore, based on homology to pea TII sequence, it appears that this peptide is missing the terminal ten residues. Although it was not possible to reach the end of the mature protein, new primer combinations allowed extending the cultivated lentil TII-like peptide of additional five residues towards the C-terminus, NSEVE, which correspond to the pea TII sequence. Experiments are ongoing to attempt isolate the complete sequence.

In order to compare at a functional level the inhibitor genes found in lentil with the behaviour of those found in pea, an analysis of the messengers produced at the level of different plant tissues was undertaken. Total RNA was isolated from immature seeds, fresh leaves and germinating roots of cultivated lentil. The cDNA obtained from total RNA was subsequently amplified using the primer combinations F1-R1 and F1-R4. Immature seed cDNA amplified only with the primer combination F1-R1, producing a fragment of the expected length; on the other hand, the cDNA from the other two tissues (leaves and roots) yielded an amplification product only with the primer combination F1-R4, while did not produce any band with F1-R1 (Fig. 4). These results indicate that *TII*-like and *TI6*-like sequences are differentially expressed in the investigated lentil organs and show a pattern of expression similar to that of the respective pea genes.



**Fig. 4** cDNA from different lentil cv. Laird organs, amplified with **a** F1-R1 and **b** F1-R4. *M* 100-bp DNA ladder

## Discussion

In all *Lens* species the two primer combinations F1-R1 and F1-R4 amplified two products differing each other for nucleotide sequence. Within each species the amplicates were expected to be similar to the *TII* and *TI6* pea genes, since the primers were designed on those sequences. A search in the data bank and the consequent FASTA alignment has confirmed such an expectation. Moreover, when all sequences were analysed using cladistic analysis, two clusters were formed, one containing all the F1-R1 generated sequences and the other including all the F1-R4 products. The same clusters included pea sequences *TII* and *TI6*, respectively. The cladogram suggests, therefore, that the F1-R1 and F1-R4 sequences are paralogous to each other in any considered species and altogether orthologous to pea *TII* and *TI6* genes, respectively.

The cladistic analysis employed the neighbour-joining method and the number of differences model. This model was chosen, since it only records differences among the sequences, correcting for missing data, without any evolutionary consideration. The cladogram so constructed merely reflects the observed differences and does not suggest any phylogenetic inference, which means that no measure of the evolutionary distances is provided. This is particularly true in the case of protein-coding DNA sequences, where the analysis of the evolutionary distances is not simplistic.

In addition, the cladistic analysis indicates that the *TII*-like sequences are highly conserved over a wide range of taxa in *Lens*, as already previously indicated (Sonnante et al. 2002), while the *TI6*-like ones are more variable, thus suggesting a possible different evolutionary rate. It has been hypothesized that trypsin-only inhibitors, like *TI6*, are more effective in protecting plants from insect pests (Hilder et al. 1989). The different pattern of expression of *TII* and *TI6* genes might have generated a distinct selective pressure on *TI6* variants, thus changing the width of the selective sieve. A further indication comes from the observation that the alpha-trypsin inhibitor (ATI) can be induced at the level of roots by soil-borne pathogens (McGurl et al. 1995). The possibly different evolutionary rate in the paralogous *BBI* coding sequence could therefore result from differences in the selective pressure operated by distinct classes of organisms at different plant organs.

Besides the great overall similarity, some residues at important positions of the reactive site loop change between *Lens* and pea *BBI*s belonging to the same class; these changes might somehow affect the functionality of *Lens* inhibitors. It has been demonstrated in synthetic polypeptides that the residue in P2' position has influence on the function of the inhibitor (Gariani et al. 1999), while the highly conserved serine in P1' is not essential for the integrity of the reactive site loop and contributes to a limited extent to the trypsin affinity. By contrast, a possible role of substitution the P5' residue

on the functionality of the inhibitor has been hypothesized (Brauer and Leatherbarrow 2003). Compared to pea, in *Lens* TI sequences we observed differences in the trypsin active sites: TI1-like proteins showed different residues in P1 and P2', while TI6-like ones displayed differences in both active sites, in P1, P1', P5' (first active site) and P2' (second active site). These substitutions may account for different specificity and biological role.

The similarity of the two types of *Lens* BBI sequences to pea BBI classes is not only observed at the level of DNA or coded protein sequence, but also suggested at the level of pattern of expression. In fact, the amplification of cDNA in cultivated lentil demonstrates that in leaves and roots the F1–R1 sequence (TI1-like) is not expressed, while the F1–R4 one (TI6-like) is expressed in the same organs; the opposite happens for the expression in immature seeds, thus showing a similar pattern of expression as observed in *Pisum* (Domoney et al. 2002).

The genes *TI1* and *TI6* in pea do not co-segregate, but belong to different linkage groups. It is therefore possible to hypothesize that also in *Lens* the two paralogous genes map on different linkage groups and possibly physically allocate on different chromosomes. Although lentil genetic maps are not well saturated (Rubeena et al. 2003; Duran et al. 2004), such a hypothesis could be verified by linkage analysis.

Cultivated lentil, *L. culinaris* subsp. *culinaris*, was domesticated from the wild taxon *L. culinaris* subsp. *orientalis*. As a consequence of domestication syndrome, two morphotypes were selected differing for seed size and sometimes for cotyledon colour, namely microsperma with small, rounder seeds and macrosperma with larger and flatter seeds (Barulina 1930; Muehlbauer et al. 1985). We have noticed that the samples of the macrosperma morphotype we analysed possess a specific autapomorphic transition A → G in the F1–R1 sequence, leading to an amino acid substitution N → D in the derived protein sequence, which is not present in all the remaining taxa of the genus. The fact that we analysed only a limited sample, constituted by two macrosperma and two microsperma, does not allow assessing with certainty that this autapomorphy is characteristic of all the macrosperma gene pool, and a much larger sample is needed to determine whether this nucleotide substitution is present in all or only in a part of this gene pool, and if it possibly has a geographic association pattern. In any case, this nucleotide substitution may be quite indicative of the evolutionary history of this morphotype and might be useful in tracing back its origin and diffusion.

A lack of homogeneity is also present in the lentil wild progenitor *L. culinaris* subsp. *orientalis*. Previous studies using different markers (van Oss et al. 1997; Ladizinsky 1999) have demonstrated that a wide variation is present in the wild progenitor of the cultigen and a comparable level of polymorphism was also detected using internal transcribed spacer sequences (Sonnante et al. 2003). Moreover, some characters are shared only by some stocks of the wild progenitor and

the cultigen. The domestication of lentil has probably taken place starting from a precise population or genetic stock; therefore, as hypothesised by Ladizinsky (1999) the analysis of such characters which are polymorphic in the wild progenitor but monomorphic in the cultigen could help identify the genetic stock from which the crop was originated. A possible example of such domestication history tracers could be the amino acidic substitution synapomorphic in the cultigen and one sample only of the wild progenitor in the TI6-like sequence in position 13 (Fig. 2).

The observation that also in *Lens* paralogous sequences coding for BBIs can be detected similarly to pea, strongly supports the idea that the presence of different classes of paralogous *TI* genes expressed at distinct levels in the plants and with different biological functions is a general rule for at least all Viciae. In an analysis of BBI sequences in flowering plants retrieved from the data banks, Mello et al. (2003) obtained a tree containing separate groups for the BBIs from dicotyledonous plants. In particular, it was evident that also in some Phaseoleae two different, potentially paralogous, groups of TI sequences were present.

On this basis we have selected some BBI sequences of the Phaseoleae and compared them with the sequences from the tribes Viciae and Trifolieae. It appears evident (Fig. 3) that, although the inhibitors from the Phaseoleae are different from those of the other Fabaceae, at least *G. max* and *V. angularis* possess two different classes of BBI protease inhibitors. This strongly supports the idea that the differentiation of this family of proteinase inhibitors at the level of sequence, function and pattern of expression observed in *Lens* and *Pisum* for BBIs coding paralogous genes is a general rule in most Fabaceae.

In the same analysis the two *M. sativa* proteinases cluster within the TI6-like group, indicating similarity to this class of inhibitors. McGurl et al. (1995) isolated these sequences from a wounded-leaf cDNA library and observed that they are expressed at the level of flowers and roots of unwounded plants too; moreover, the active sites of the alfalfa inhibitors are both trypsin inhibitors like pea TI6 and TI9 proteins.

In *Pisum* TI6 and TI9 sequences, both double-trypsin inhibitors, are more similar to each other than to TI1, indicating a more-recent evolution. It could be possible that in taxa other than *Pisum*, the differentiation observed in pea has not occurred to the same extent and therefore the double-trypsin-inhibitor sequences retain similarity to both TI6 and TI9 of pea. This is also evident in Fig. 3, where the branching of the TI6-like group is supported by a high bootstrap value, but sub-branching between TI6 and TI9 is not sustained.

Taken altogether, these data strongly indicate that at least two different genes coding for BBI protease inhibitors are present in all Fabaceae, and that possibly such an organisation was present in the progenitor of the whole family.



**Acknowledgements** We thank the ICARDA (Syria) and USDA (USA) genebanks that donated the seeds. This research was partially funded by MiUR, Programma Operativo del Piano 'Ingegneria molecolare' - Progetto n. 2-'Geni d'interesse biomedico ed agroalimentare'.

## References

- Ahmad M, McNeil DL, Sedcole JR (1997) Phylogenetic relationships in *Lens* species and their interspecific hybrids as measured by morphological characters. *Euphytica* 94:101–111
- Barulina HI (1930) Lentils of the USSR and other countries. *Bull Appl Bot Plant Breed (Leningrad)* 40 [Suppl]:1–319
- Brauer ABE, Leatherbarrow RJ (2003) The conserved P1' Ser of Bowman–Birk-type proteinase inhibitors is not essential for the integrity of the reactive site loop. *Biochem Biophys Res Comm* 308:300–305
- Brauer ABE, Kelly G, McBride J, Cooke RM, Matthews AJ, Leatherbarrow RJ (2001) The Bowman–Birk inhibitor reactive site loop sequence represents an independent structural  $\beta$ -hairpin motif. *J Mol Biol* 306:799–807
- Clemente A, MacKenzie DA, Jeenes DJ, Domoney C (2004) The effect of variation within inhibitory domains on the activity of pea protease inhibitors from the Bowman–Birk class. *Protein Expr Purif* 36:106–114
- Domoney C, Welham T, Sidebottom C, Firmin JL (1995) Multiple isoforms of *Pisum* trypsin inhibitors result from modification of two primary gene products. *FEBS Lett* 360:15–20
- Domoney C, Welham T, Ellis N, Mozzanega P, Turner L (2002) Three classes of proteinase inhibitor gene have distinct but overlapping patterns of expression in *Pisum sativum* plants. *Plant Mol Biol* 48:319–329
- Durán Y, Fratini R, García P, Pérez de la Vega M (2004) An intersubspecific genetic map of *Lens*. *Theor Appl Genet* 108:1265–1273
- Ferguson ME, Maxted N, Slageren M. van, Robertson LD (2000) A re-assessment of the taxonomy of *Lens* Mill. (Leguminosae, Papilionoideae, Viciae). *Bot J Linn Soc* 133:41–59
- Gariani T, McBride JD, Leatherbarrow RJ (1999) The role of P2' position of Bowman–Birk proteinase inhibitor in the inhibition of trypsin. Studies on P2' variation on cyclic peptides encompassing the reactive site loop. *Biochim Biophys Acta* 1431:232–237
- Havey MJ, Muehlbauer FJ (1989) Variability for restriction fragment lengths and phylogenies in lentil. *Theor Appl Genet* 77:839–843
- Hilder VA, Barker RF, Samour RA, Gatehouse AMR, Gatehouse JA, Boulter D (1989) Protein and cDNA sequences of Bowman–Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp.). *Plant Mol Biol* 13:701–710
- Hoffman DL, Soltis DE, Muehlbauer FJ, Ladizinsky G (1986) Isozyme polymorphism in *Lens* (Leguminosae). *Syst Bot* 11:392–402
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Ladizinsky G (1999) Identification of lentil's wild genetic stock. *Genet Res Crop Evol* 46:115–118
- Ladizinsky G, Braun D, Goshen D, Muehlbauer FJ (1984) The biological species of the genus *Lens* L. *Bot Gaz* 154:253–261
- Maki PA, Paterson Y, Kennedy AR (1994) Studies related to the potential antigenicity of the Bowman–Birk inhibitor, an anti-carcinogenic protease inhibitor isolated from soybeans. *Nutr Cancer* 22:185–193
- Mayer MS, Soltis PS (1994) Chloroplast DNA phylogeny of *Lens* (Leguminosae): origin and diversity of the cultivated lentil. *Theor Appl Genet* 87:773–781
- McGurl B, Mukherjee SK, Kahn ML, Ryan CA (1995) Cloning and characterization of two Bowman–Birk proteinase inhibitors from alfalfa (*Medicago sativa* var. vernema). *Plant Mol Biol* 27:995–1001
- Mello MO, Tanaka AS, Silva-Filho MC (2003) Molecular evolution of Bowman–Birk type proteinase inhibitors in flowering plants. *Mol Phylogenet Evol* 27:103–112
- Muehlbauer FJ, Cubero JI, Summerfield RJ (1985) Lentil (*Lens culinaris* Medic). In: RJ Summerfield, EH Roberts (eds) Grain legume crops. Collins, London, pp 266–311
- Muench DG, Slinkard AE, Scoles GJ (1991) Determination of genetic variation and taxonomy in lentil (*Lens* Miller) species by chloroplast DNA polymorphism. *Euphytica* 56:213–218
- Odani S, Ikenaka T (1976) The amino acid sequence of two soybean double headed proteinase inhibitors and evolutionary considerations on the legume proteinase inhibitors. *J Biochem* 80:641–643
- Pinkas R, Zamir D, Ladizinsky G (1985) Allozyme divergence and evolution in the genus *Lens*. *Plant Sys Evol* 151:131–140
- Prakash B, Selvaraj S, Murthy MRN, Sreerama YN, Rao DR, Gowda LR (1996) Analysis of the amino acid sequences of the plant Bowman–Birk inhibitors. *J Mol Evol* 42:560–569
- Rohrmeier T, Lehle L (1993) WIP1, a wound inducible gene from maize with homology to Bowman–Birk proteinase inhibitors. *Plant Mol Biol* 22:783–792
- Rubeena RF, Taylor PWJ (2003) Construction of an intraspecific linkage map of lentil (*L. culinaris* ssp. *culinaris*). *Theor Appl Genet* 107:910–916
- Ryan CA (1990) Protease inhibitors in plants: genes for improving defences against insects and pathogens. *Annu Rev Phytopathol* 28:425–449
- Sharma SK, Knox MR, Ellis THN (1996) AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. *Theor Appl Genet* 93:751–758
- Small RL, Wendel JF (2002) Differential evolutionary dynamics of duplicated paralogous *Adh* loci in allotetraploid cotton (*Gossypium*). *Mol Biol Evol* 19:597–607
- Sonnante G, De Paolis A, Pignone D (2002) Trypsin inhibitor gene sequence analysis in the genus *Lens* Mill. International congress on molecular evolution, Sorrento, 13–16 June 2002, p 681
- Sonnante G, Galasso I, Pignone D (2003) ITS sequence analysis and phylogenetic inference in the genus *Lens* Mill. *Ann Bot* 91:49–54
- Van Oss H, Aron Y, Ladizinsky G (1997) Chloroplast DNA variation and evolution in the genus *Lens* Mill. *Theor Appl Genet* 94:452–457