1-1-2004

TCA Microsatellite Repeats in the 5'UTR of the Sat5 Gene of Wild and Cultivated Accessions of Pisum and of Four Closely Related Genera

Joel P. Stafstrom

Paul Ingram

Original Citation
Stafstrom, J.P., and Ingram, P. 2004. TCA microsatellite repeats in the 5'UTR of the Sat5 gene of wild and cultivated accessions of Pisum and of four closely related genera. Int. J. Plant Sci. 165:273-280.

Int. J. Plant Sci. 165(2):273–280. 2004.

1058-5893/2004/16502-0005$15.00

TCA MICROSATELLITE REPEATS IN THE 5′UTR OF THE Sat5 GENE OF WILD AND CULTIVATED ACCESSIONS OF *PISUM* AND OF FOUR CLOSELY RELATED GENERA[1]

Joel P. Stafstrom[2] and Paul Ingram

Plant Molecular Biology Center, Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115, U.S.A.

PsSat5, a cDNA clone from *Pisum sativum* cv. Alaska, contained a microsatellite consisting of 15 TCA repeats within the 5′UTR. This SSR microsatellite was immediately upstream of the presumptive ATG start codon. PCR amplification of genomic DNA from cv. Alaska yielded an identical sequence. This repeat region was analyzed from 10 additional wild and cultivated accessions of *Pisum* and from four closely related genera (*Cicer*, *Lathyrus*, *Lens*, and *Vicia*). All of the sequences were generally quite similar, with the exception of the number of TCA repeats (region 3) and a short domain immediately upstream of the repeats (region 2). *Pisum humile*-northern and *Lathyrus* each contained four TCA repeats (the fewest number observed). Similar to *P. sativum*-Alaska and other cultivated peas, *Lens* contained 15 repeats, the largest number observed. The number of TCA repeats does not appear to correspond to the established phylogeny of these accessions, so the cellular events that generated variable numbers of repeats probably have occurred repeatedly and have involved both expansions and contractions in the number of repeats. The mRNA corresponding to PsSat5 was found in all tissues of *P. sativum*-Alaska that were examined, but its abundance in leaves and sepals was low. The level of expression was similar in growing and nongrowing stems, roots, and axillary buds. Northern blot analysis of stems and leaves of all 15 accessions showed similar levels of expression. Therefore, there is not a clear correlation between the number of TCA repeats in the 5′UTR and the level of Sat5 expression.

Keywords: microsatellite, *Pisum*, SSR, 5′ untranslated region, 5′UTR, Vicieae.

Introduction

Eukaryotic genomes, especially large genomes, are rich in repetitive DNA. Microsatellites, one class of repetitive DNA, consist of short DNA sequences (up to about six nucleotides) that occur in variable numbers of tandem repeats (Tautz 1989). Microsatellites also are referred to as SSRs, simple sequence length polymorphisms (SSLPs), and short tandem repeats (STRs). Several classes of SSRs appear to be common in the genomes of all plants (Morgante and Olivieri 1993; Wang et al. 1994). The number of repeat elements at a locus may be quite variable. For example, 37 alleles of the barley *Waxy* gene were identified based on the number of AT repeats within an intron of this gene (Saghai Maroof et al. 1994). SSR polymorphisms mutate and evolve rapidly, which opens up the possibility of detecting allelic differences between closely related species, within a species, or even among individuals in a population. SSR markers have been used to construct genetic maps for several species, including *Arabidopsis* (Bell and Ecker 1994), *Phaseolus* and *Vigna* (Yu et al. 1999), *Vitis* (Sefc et al. 1999), and *Cicer* (Hüttel et al. 1999; Udupa et al. 1999). In addition, SSR polymorphisms are useful for fingerprinting individuals and their progeny in population and ecological studies (Innan et al. 1997; Li et al. 2000). SSRs also are being used for varietal identification, for example, in Basmati rice breeding programs (Nagaraju et al. 2002). SSRs commonly occur in intergenic regions, but they may also occur in various regions of genes. SSRs are frequently presumed to be selectively neutral. However, certain SSR alleles in wild barley were correlated with growth in various microniches (Huang et al. 2002). In this plant, it was not clear whether these SSRs were direct targets of selection or whether they were markers for haplotypes that were under selection.

Several human neurological disorders are caused by SSRs within various regions of the affected gene. Huntington's Disease is a dominant neurological disorder caused by a trinucleotide repeat within the coding region of the gene (Reddy and Housman 1997). The normal gene contains 6–34 CAG repeats (encoding a polyglutamine tract), whereas disease symptoms appear and become progressively more severe when 36–121 CAG repeats are present. Spinal and bulbar muscular atrophy, Machado-Joseph disease, and several spinalcerebellar ataxias also are the result of expanded CAG repeat regions within the coding region of the corresponding gene (Reddy and Housman 1997). Trinucleotide repeats and other SSRs within noncoding regions of genes cause other human diseases. For example, CTG repeats within the 3′UTR of the myotonin gene lead to myotonic dystrophia, GAA repeats in an intron of

[1] Sequences reported in this article are deposited in the GenBank database under the following accession numbers: PsSat5 cDNA, AY367058; 15 Sat5 genomic fragments, AY370636 to AY370650. Abbreviations: SNP, single nucleotide polymorphism; SSR, simple sequence repeat; UTR, untranslated region.

[2] Author for correspondence; e-mail stafstrom@niu.edu.

Manuscript received August 2003; revised manuscript received December 2003.

the frataxin gene cause Friedreich's ataxia, and CGG repeats within the 5′UTR of the FMR-1 gene are associated with fragile X syndrome (Reddy and Housman 1997).

SSRs have been found within the 5′UTRs of a number of plant and animal genes. In some organisms, these polymorphisms are correlated with altered patterns of gene expression. In rice, a variable number of CT repeats in the 5′UTR of the *Waxy* gene is correlated with amylose content in grains. However, most of the variability in amylose content is the result of an SNP adjacent to this microsatellite, which appears to affect an intron splice site (Bligh et al. 1995; Ayres et al. 1997; Larkin and Park 1999). The pea ferredoxin-1 gene is more strongly expressed in light than in darkness (Dickey et al. 1998). This differential expression is in part the result of a short microsatellite within the 5′UTR, $CAAT_4$, which appears to render ribosome-associated mRNAs less susceptible to degradation. Translational regulation by light of spinach *PsaD* also involves its 5′UTR, which contains several CAA repeats (Sherameti et al. 2002). In this case, however, the role of the repeats per se in controlling translation was not established. The presence of a microsatellite within the 5′UTR does not necessarily imply that polymorphic alleles will exist in the population. For example, the human MAP-2 gene contains seven CAG repeats in its 5′UTR. The same number of repeats was found in a total of 86 individuals, including 31 controls and 55 individuals with neuropsychiatric illnesses (Kalcheva et al. 1999). The human calmodulin-1 gene also contains seven CAG repeats in its 5′UTR (Toutenhoofd et al. 1998). Again, no variability is known to occur in the population. However, experimental manipulation of the number of CAG repeats affected gene expression: elimination of these repeats reduced expression by 45%, whereas replacement with 20 or 45 CAG repeats had no effect on the level of expression (Toutenhoofd et al. 1998). A final example highlights the importance of the particular repeat sequence. Increasing numbers of either CAG or CTG repeats within the 5′UTR of a human marker gene reduced the rate of translation (Raca et al. 2000), but the reduction was about twofold greater for CTG repeats than for CAG repeats. It was hypothesized that CUG repeats in single-stranded mRNA would form more stable hairpins and thereby attenuate translation to a greater degree.

The haploid genome of *Pisum sativum* contains about 4.43 pg of DNA (Bennett and Leitch 1997), ca. 97% of which is estimated to be repetitive DNA (Murray et al. 1981). Several families of highly abundant tandem and dispersed repeats have been identified in *Pisum* (Neumann et al. 2001). In the course of screening for growing-bud associated cDNAs (Devitt and Stafstrom 1995), we isolated a clone from pea (*P. sativum* L. cv. Alaska) that contained an SSR consisting of 15 TCA repeats immediately upstream of the putative ATG start codon. We refer to this gene as PsSat5 to indicate that this microsatellite is in the 5′UTR. In this work, we addressed three basic questions. Is there variability at this locus in other accessions of *Pisum* and in closely related species? If so, is the amount or type of variability correlated with the established phylogeny of these plants? And finally, is sequence variability at this locus correlated with variable expression of the Sat5 gene?

Material and Methods

Plants

Eleven accessions of *Pisum* and four related genera were studied. *Pisum sativum* cv. Alaska was purchased from W. Atlee Burpee (Warminster, Pa). The remaining accessions were obtained from Dr. Neil Polans (Northern Illinois University, DeKalb). Accessions of *P. sativum* were cv. Parvus (L1107); tester line A7391; and isolates from Ethiopia (JI171), Greece (JI264), and Turkey (JI1035). Other *Pisum* species were *Pisum abyssinicum* (JI227); *Pisum elatius* (isolate 721); *Pisum fulvum* (isolate 701); *Pisum humile*-northern (isolate 716); and *Pisum humile*-southern (isolate 713). Other genera studied were *Cicer arietinum* cv. Sonora 80; *Lathyrus sativus* cv. P-534; *Lens culinaris* cv. Chilean 78; and *Vicia faba* cv. Mammoth. Seeds were scarified, sown in Promix, and grown in a greenhouse or in a growth chamber. Ambient temperature in the greenhouse ranged from ca. 20° to 30°C, depending on the season. Halogen lamps were used to extend photoperiods to 16 h during all seasons. Growth chamber temperature was set at 25°C. Mixed fluorescent and incandescent illumination was provided over a 16L:8D photoperiod.

Isolation of the PsSat5 cDNA Clone

The PsSat5 clone was isolated from a growing-bud cDNA library prepared from cv. Alaska (Devitt and Stafstrom 1995). In order to fully sequence both strands of this clone, nested deletion clones were generated using the Erase-A-Base kit (Promega). Subclones were sequenced either manually (Sequenase II kit, US Biochemical, Cleveland, Ohio) or using an Applied Biosystems PRISM 373 automated sequencer (Biotechnology Core Facility, Northern Illinois University, DeKalb).

PCR Amplification and Sequencing of Sat5 Gene Fragments

Genomic DNA was isolated using the Nucleon Phytopure DNA extraction kit (Amersham). Oligonucleotide primers based on the PsSat5 cDNA clone were used to amplify the corresponding region by PCR. The sequence of the forward primer (202) was 5′-ttctAGACCAACGCGGAGGG, and the sequence of the reverse primer (203) was 5′-ggaatTCCTGTGTTGGTTGTTCAGC (the bases shown in lowercase were added to create Xba1 and EcoR1 restriction sites, respectively). PCR amplification was done using Taq polymerase and 35 cycles of the following protocol: 94° for 1 min; 55° for 2 min; and 72° for 3 min, plus a 5-s extension for each cycle. PCR products were gel-purified, digested with Xba1 and EcoR1, and cloned into pSPORT1 (Gibco-BRL). Clones were obtained for all templates except *Vicia*. Both strands were sequenced using SP6 and T7 primers contained in the vector. To verify the results of this analysis, DNA was extracted from a second set of plants and amplified by PCR. In this case, the PCR products were gel-purified and sequenced directly. Both strands were sequenced using primers 202 or 203.

RNA Gel Blotting

Several tissues and organs from *P. sativum*-Alaska were analyzed. Growing stem tissue was isolated from internode-5 of 10-d-old pea plants. Mature stem tissue was from internode-3 of 10-d-old plants. Leaflets were from leaves at node-3 or node-4 of 10-d-old plants. Floral organs (sepals, petals, stamens, and carpels) were from flowers 1 d before pollen anthesis, a stage that was easy to recognize and occurred ca. 21 d after sowing under these growth conditions. Root apices (terminal 2 mm, including root cap) were isolated from 3-d-old seedlings sown on moist paper towels. Fully elongated root tissue (10–20 mm from the apex) was collected from the same seedlings. In addition, leaves and mature stems were collected from all 15 accessions. Isolated organs were frozen in liquid nitrogen and stored at −80°C. Total RNA was extracted using guanidinium isothiocyanate and phenol followed by LiCl precipitation (Stafstrom and Sussex 1992). RNA (10 μg/lane) was separated by denaturing formaldehyde gel electrophoresis and was blotted onto nylon membranes by capillary transfer. Random-primed ^{32}P-labeled probes were prepared using inserts from the corresponding cDNA clones as templates (DecaPrime II kit, Ambion). Blots were prehybridized in 50% formamide, 5X Denhardt's reagent, 0.5% SDS, 5X SSPE, and 100 μg/mL salmon sperm DNA for 2 h at 42°C and hybridized overnight under the same conditions. Blots were washed in 0.2X SSC and 0.1% SDS at 55°C. Data were collected on x-ray film or using a Storm 820 Phosphorimager (Amersham/Molecular Dynamics). Blots were stripped and reprobed up to two times. RNA loadings were assessed by ethidium bromide staining of rRNAs or by probing blots with BB695, which recognizes 28S rRNA (gift of L. Kaufman, University of Illinois, Chicago).

Results

A pea axillary bud cDNA library (*Pisum sativum* L. cv. Alaska) was screened for clones that were more highly expressed in growing buds than in dormant buds. One putative growth-associated clone contained 15 repeats of the trinucleotide TCA (fig. 1). The first in-frame ATG codon occurred immediately after the TCA repeats. Although this cDNA did not contain an in-frame stop codon upstream of this ATG, other accessions contained such stop codons (see below). We refer to this cDNA as PsSat5 to indicate that the TCA microsatellite occurs within the 5′UTR. The fully sequenced *Arabidopsis thaliana* and *Oryza sativa* genomes each contain one Sat5-related gene (At2G37110 and Gb BAB08185, respectively). The deduced protein sequences ranged from 235 to 242 residues and contained a DUF614 domain. Sequence identity between PsSat5 and *Arabidopsis* and *Oryza* Sat5 proteins was 68% and 59%, respectively (data not shown). PsSat5-related sequences were identified in all plant genomes that have been studied in detail. *Medicago truncatula* EST clones (e.g., BG580217) contained two TCA repeats immediately before the ATG start codon, whereas *Glycine max* EST clones (e.g., BG046679) contained a single TCA unit in this position. SSRs were not found in the 5′UTRs of Sat5 genes from *Arabidopsis*, *Oryza*, or *Zea* (data not shown).

Genomic DNA was isolated from 11 *Pisum* accessions (*P. sativum*, *Pisum abyssinicum*, *Pisum elatius*, *Pisum fulvum*,

```
cagcattaca ttgcatagaa gagaagacca acgcggaggg attttggcct   50
ccaccgacat ctcttctcaa ttctcaaatt cacatcaatt atatttctca  100
tcatcatcat catcatcatc atcatcatca tcatcatcat caATGGCTGA  150
TGATAACAAG AAACTCAATC ACGACGAGGA GGCAAATCCT CTTCTCCACC  200
ACCACCATCA CCAACAACAA CCACCACCTA AAGCAGCTGA ACAACCAACA  250
CAGGAGGAAC TTTTTCATCT ATGGACCGCT GATGGTGTTC CCATCGCCCA  300
TGGAAGCGTA ATGGGCCAGC CTATTCCCCG TTCTCCTTGG AATTCCAGCG  350
TCTGCGCTTG TCTCGGTCAA AGCGATCACT TTTGCAGCAG CGATCTCGAA  400
GTTTGTCTTC TTGGGAGTGT GGCTCCTTGT GTGCTGTATG GAAGCAATGT  450
TGAGAGACTT GGGTCCAATA ATTCCGGGAC ATTTGCCAAT CATTGTTTGC  500
ATTATTCTGG CCTCTATGTA ATTGGGAATT CCTGCTGTGG TTGGAATTGT  550
CTTGCGCCGT GGTTTTCATA TCCTAGCCGA ACTGCAATTC GTCGCAGGTT  600
CAATTTGGAG GGAAGCTGTG AGGCACTTAA TAGGTCATGT GGGTGCTGCG  650
GAAGCTTTTT GGAAGATGAG GCGCAGCGTG AACAGTGCGA ATTGGCATGT  700
GACTTTGCAA CCCATTTCTT TTGTCATGCC TGTGCTCTTT GTCAAGAAGG  750
TCGTGAGCTC CGCCGTAGAG TGCCTCATCC TGGCTTCAAT GCGCAACAAA  800
TATTGGTTAT GATCCCACCT GCAGAGCAGG CAATGGGTCG TGGGGCGTGA  850
aaaacattct tcctcatctt ggaagtgcta ttctggtgta tgctattact  900
gttgctttga tatgatttga cacgtcgaac tgctgcctcc cttttctcca  950
gattttgctg aagagtatca tggttgtttg tttgagtttg gtataaataa 1000
agaacataca gcaccattat attttcagtt tgtatcctct actattctgt 1050
agtatatatg tattttgttg caatatgaat ttatatgtat tattgagggc 1100
atggttcctg tatggttatg agtggttgta gcctgtaatt cagtgaatgc 1150
atattatgtt ttaaatataa aaaaatagtt aatatccaaa aaaaaaaaaa 1200
aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 1250
aaaaaaaaaa aaaaaaa                                     1367
```

Fig. 1 Nucleotide sequence of PsSat5. The cDNA consisted of 1367 bp, including a 705-bp deduced coding region (uppercase). The 5′UTR contained a microsatellite that consisted of 15 TCA repeats (underlined). The sequences used to construct forward and reverse PCR primers flanking the TCA repeat region are double-underlined (primers 202 and 203, respectively). GenBank accession AY367058.

Pisum humile), from *Lens*, *Lathyrus*, and *Vicia* (tribe Vicieae), and from *Cicer* (tribe Cicerae). Each DNA was used as a template for PCR amplification using primers that flank the TCA-repeat region in PsSat5 (see fig. 1). Based on the PsSat5 cDNA, the expected DNA fragment size was 239 bp. Each *Pisum* accession gave rise to a major PCR product that migrated between the 220 and 298 bp markers (fig. 2). The four other genera gave rise to major bands of similar size and additional bands. These additional bands are presumed to be PCR artifacts.

For each accession (except *Vicia*), the major PCR band was cloned into a plasmid vector and sequenced. To confirm the sequences obtained from these clones, DNA was isolated from a second set of plants and amplified by PCR. In this experiment, the PCR products from all 15 accessions were isolated from agarose gels and sequenced directly. Sequences obtained by each method were identical, with two minor exceptions: for *P. abyssinicum*, one sequence contained nine TCA repeats and the other contained 10 repeats, and for

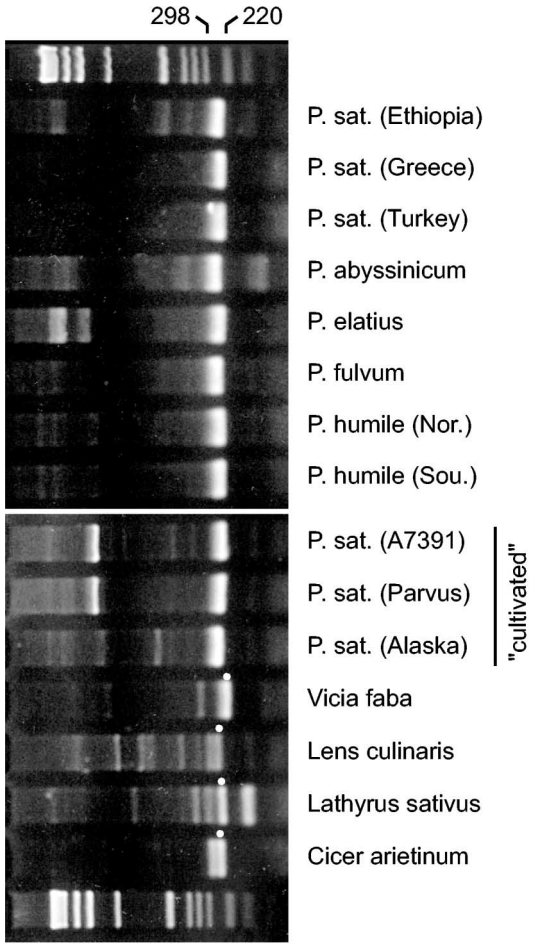

Fig. 2 PCR amplification of genomic DNA from 11 *Pisum* accessions and from four related genera. A major DNA fragment from all *Pisum* accessions migrated between the 220- and 298-bp markers. This fragment was cloned and sequenced (experiment 1) or sequenced directly (experiment 2). For each of the other genera, the marked band was isolated.

P. sativum-Ethiopia, codon 19 corresponded to either histidine (CAC) or leucine (CTC). The nucleotide sequence of the PsSat5 cDNA was compared with the sequences of the 15 PCR fragments (fig. 3). PCR products from three "cultivated" pea accessions were identical (Alaska, Parvus, and A7391). Excluding the PCR primers, the PCR fragment from cultivated peas consisted of 193 bp. Four regions were identified. Region 1 was quite similar among all of the accessions, despite the presence of a few SNPs (several in *Vicia*) and short deletions in *P. abyssinicum* and *Vicia*. Region 2, which was absent in cultivated peas, contained 0–25 bases. Five of the accessions contained stop codons in the same reading frame as the putative initiator ATG codon. Region 3 included the TCA repeats. The number of TCA repeats varied from a minimum of four (*P. humile*-northern, *Lathyrus*) to a maximum of 15 (cultivated *P. sativum*, *Lens*). As mentioned above, one *P. abyssinicum* sequence contained nine TCA repeats and the other contained 10. Region 4 corresponds to the first 31 codons of the coding region. *Lathyrus*, *P. fulvum*, and *Vicia* contained deletions corresponding to two, four, and six codons, respectively, which encoded histidine or asparagine residues. The total size of region 2 plus region 3 ranged from 32 to 45 bps. The sequence of cultivated *P. sativum* was identical to that of *Lens* and differed from that of *Cicer* only in the number of TCA repeats (15 rather than 14).

PsSat5 expression in cv. Alaska was analyzed by Northern blotting (fig. 4). A single 1.4 kb messenger RNA was detected. Growing roots (root apices), mature roots, elongating stems, and mature stems contained similar amounts of this mRNA. The amount of message in leaves was relatively low. Expression in floral organs was also examined. Relatively high levels of PsSat5 mRNA were present in stamens; petals and carpels contained moderate amounts, but little was detected in sepals. Dormant and growing axillary buds contained similar amounts of PsSat5 message (data not shown). The overall pattern of PsSat5 expression is distinct from that of histone H2A, a growth-associated marker gene.

Accumulation of PsSat mRNA in leaves and mature stems of all 15 accessions was examined (fig. 5). A single band was seen in each lane. Similar to results with *P. sativum*-Alaska (fig. 4), the level of message accumulation was typically two- to threefold higher in stems than in leaves. Slight differences are probably from loading errors (e.g., *P. humile*-northern stems were under-loaded, *P. sativum*-Greece leaves were over-loaded). After accounting for loading differences, *Vicia*, *Lens*, and all of the *Pisum* accessions contained rather similar message levels in each tissue. In contrast, mRNA levels in stems and leaves of *Lathyrus* and *Cicer* appear to be reduced relative to the other accessions.

Discussion

PsSat5 was isolated from a *Pisum sativum*-Alaska growing-bud cDNA library. The most striking feature of this clone was the presence of an SSR microsatellite immediately upstream of the ATG start codon. This SSR consisted of 15 tandem TCA repeats (fig. 1). It was hypothesized that this region would be variable among plants related to *P. sativum*-Alaska and that the level and type of variability would reflect the

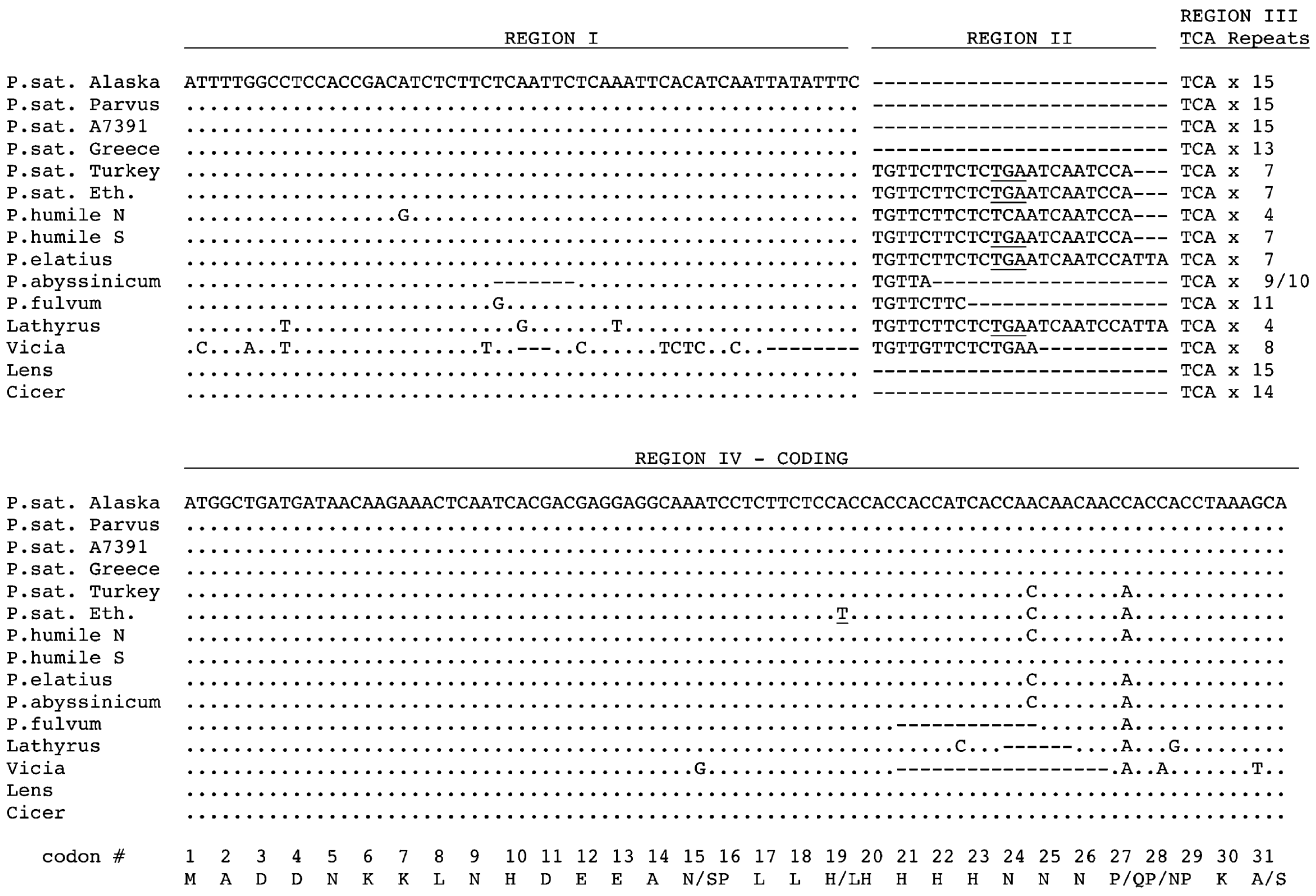

Fig. 3 Comparison of nucleotide sequences from 11 *Pisum* accessions and from four related genera. Taxa are arranged (top to bottom) according to the established phylogeny. Nucleotide identity with cv. Alaska is indicated by a dot; gaps are indicated by dashes. In-frame stop codons are underlined. Sequences for each accession were determined twice (see "Material and Methods" for details). Only two discrepancies were seen. For *Pisum abyssinicum*, nine TCA repeats were present in one sequence and 10 in the other. For *Pisum sativum*-Ethiopia, codon 19 was CAC (His) in one case and CTC (Leu) in the other. Sequences corresponding to the primers were identical in all cases and are not shown.

established phylogeny of these plants. Furthermore, sequence variability within the 5′UTR might influence gene expression at the level of Sat5 mRNA accumulation.

Ca. 200 bp of genomic DNA flanking the TCA repeat region was analyzed from 11 accessions of *Pisum* and from *Lens*, *Vicia*, *Lathyrus*, and *Cicer* (fig. 3). The set of plants used here was based on a collection used in a previous study to infer a phylogeny for the genus *Pisum* based on RAPD, allozyme, and morphological markers (Hoey et al. 1996). *Pisum sativum* included three common cultivars (Alaska, Parvus, and tester line A7391) plus accessions from Greece, Turkey, and Ethiopia. *Pisum humile*-northern and *Pisum elatius* had been shown to be more closely related to *P. sativum* than *P. humile*-southern. *Pisum fulvum* is commonly considered to be a distinct species. Recent and more extensive analyses of *Pisum* phylogeny have utilized AFLP markers and *PDR1* retro-element insertion site markers (Ellis et al. 1998; Pearce et al. 2000; Knox and Ellis 2001). From these studies, *P. fulvum* and *Pisum abyssinicum* emerged as clear groups, supporting their designation as distinct species. A third group included *P. sativum* cultivars and wild accessions as well as *P. humile*, *P. elatius*, and *Pisum jomardii*. Within this last group, most *P. elatius* accessions clustered together as did most *P. sativum* accessions. In addition, we examined four other genera. *Pisum*, *Lens*, *Vicia*, and *Lathyrus* all represented tribe Vicieae, whereas *Cicer* represented tribe Cicereae (Polhill 1981). On the basis of these phylogenetic studies, the Sat5 sequences are listed in order of increasing phylogenetic distance from cultivated pea (fig. 3). A progressive increase in sequence variability is seen between *P. sativum*-Alaska (region 2 absent, 15 TCA repeats) and *P. humile*-northern (region 2 present, four TCA repeats). However, these trends reverse as the phylogeny progresses through other *Pisum* taxa. In *P. abyssinicum* and *P. fulvum*, which are considered to be distinct species, the size of region 2 is reduced to 5–8 bp, and the number of TCA repeats is increased to between nine and 11. Thus, this 200-bp genomic region of Sat5 genes does not faithfully predict phylogeny within *Pisum*. This conclusion implies that regions 2 and 3, but not regions 1 and 4, which flank them, have undergone relatively rapid change. Interestingly, one isolate of the *P. abyssinicum* contained nine TCA repeats and the other contained 10 repeats (fig. 3), indicating that variability can occur even within small populations of closely related individuals.

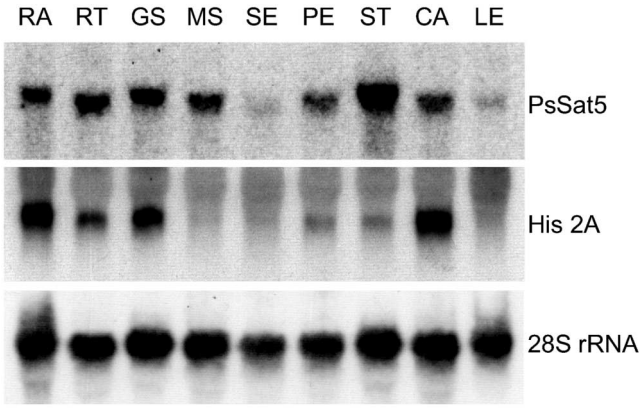

Fig. 4 Northern blot analysis of gene expression in Alaska pea plants. The PsSat5 probe recognized a single message of ≈1400 bases. Similar amounts of this mRNA occurred in root apices (*RA*), fully elongated roots (*RT*), growing stems (*GS*), mature stems (*MS*), petals (*P*), and carpels (*C*). Stamens (*ST*) contained relatively high levels of this mRNA, whereas message accumulation in sepals (*SE*) and leaves (*LE*) was quite low. This blot was stripped and reprobed with histone H2A, a growth-associated marker gene, and finally with a rRNA probe as a loading control. Total RNA (10 μg/lane) was separated by denaturing formaldehyde gel electrophoresis and transferred to a nylon membrane. The membrane was hybridized with ^{32}P oligo-labeled probes.

The greatest amount of variability occurred in regions 2 and 3 (fig. 3). The size of region 2 ranged from 0 to 25 bp, whereas the size of region 3 ranged from 12 to 45 bp (four to 15 TCA repeats). Despite this variability, the overall size of region 2 plus region 3 was rather similar, ranging from 32 bp (*P. abyssinicum*) to 45 bp (cultivated *P. sativum*, *Lens*). Thus, there was an inverse correlation between the sizes of these regions: as one region expands, the other would appear to contract. Such a linkage would lead to a relatively constant spacing between the start of translation and putative upstream regulatory elements. Variability in the number of TCA repeats may be the result of slipped-strand mispairing during DNA replication (Djian 1998); possible mechanisms leading to changes in region 2 are less clear.

The sequences of cultivated *P. sativum* and of two distantly related taxa, *Lens* and *Cicer*, were nearly identical across the entire 200-bp region. The only discrepancy was that *Cicer* contained 14 rather than 15 TCA repeats. With the exception of *Vicia*, all of the accessions were analyzed twice, beginning with DNA isolated from different plants. Therefore, it seems unlikely that these unexpected results were due to contamination during PCR or mislabeling of samples.

The cellular function of PsSat5 is not well understood. PsSat5 was isolated during a screen from growing-bud associated clones. However, expression levels in growing compared with nongrowing stems and roots (fig. 4) and axillary buds (data not shown) were similar. Also, the PsSat5 expression pattern was not similar to that of histone H2A, a growth-associated marker gene (fig. 4; Devitt and Stafstrom 1995; Stafstrom et al. 1998). The deduced protein encoded by PsSat5 would contain 235 amino acid residues. A Phi/Psi BlastP search revealed the presence of a DUF614 domain (protein family pfam04749). This domain has been conserved in eukaryotic organisms including higher plants, green algae, mammals, and euglenoids, but the function of this domain is not known.

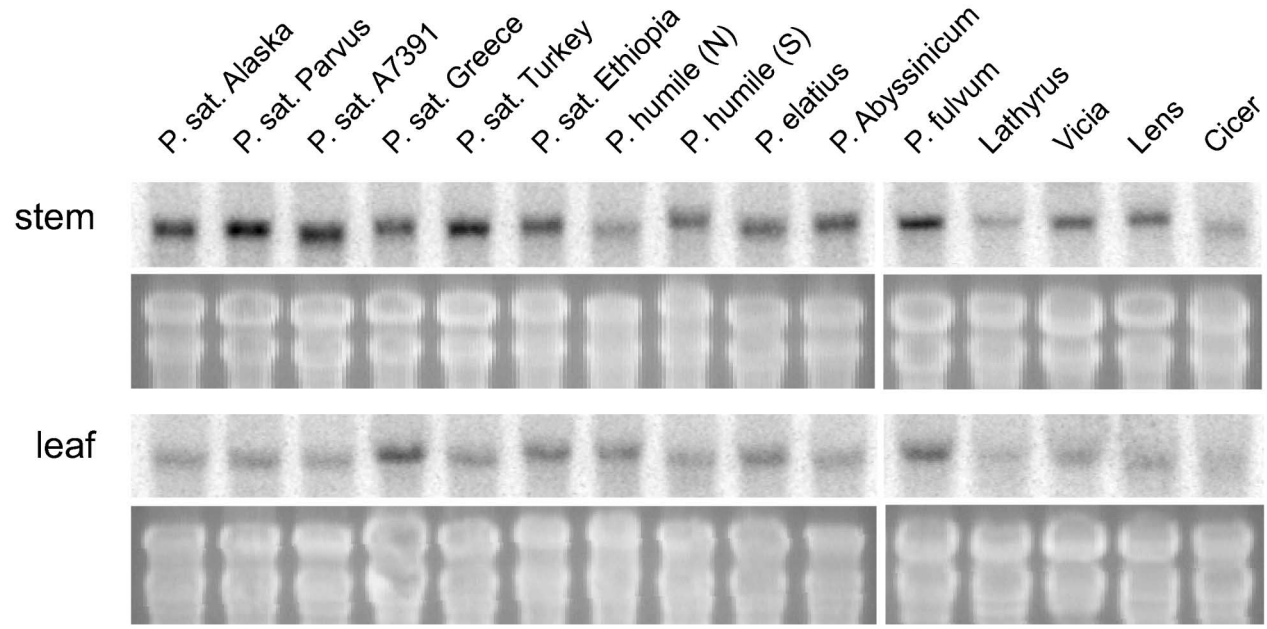

Fig. 5 Northern blot analysis of leaf and mature stem tissues. In general, the amount of mRNA in each tissue was similar for all accessions. Small discrepancies appear to be due to slight differences in loading, as assayed by ethidium bromide staining of rRNAs. Methods were as for figure 4.

SSRs have been found within the 5′UTRs of a number of plant and animal genes. In some organisms, variability in the number of repeats, either naturally occurring or experimentally induced, has been correlated with increases or decreases in gene expression (Dickey et al. 1998; Toutenhoofd et al. 1998; Raca et al. 2000). There was not much difference in Sat5 mRNA accumulation among the accessions studied here (fig. 5). Expression in both *Lathyrus* and *Cicer* appeared to be reduced relative to the other accessions. Because *Lathyrus* had four TCA repeats and *Cicer* had 14 repeats, reduced PsSat5 expression does not appear to be correlated with the number of repeats. It is more likely that promoter elements outside of the 5′UTR account for these small differences in mRNA abundance. It remains to be determined if TCA repeats in the Sat5 gene can influence translation. In none of the instances cited above did an SSR occur immediately before the ATG start codon. It has been suggested that hairpin formation within the 5′UTR of the mRNA can affect translation (Raca et al. 2000). The longest repeat we found in PsSat5 consisted of 15 TCA units. On the basis of thermodynamic calculations, a UCA_{15} tract in PsSat5 mRNA would not form stable hairpins at normal growth temperatures (Mathews et al. 1999).

Acknowledgments

We thank Neil Polans for suggesting the plant lines used in this study, for providing seed, and for helpful discussions; Michelle Devitt for help with Northern blotting; and Scott Grayburn for help with DNA sequencing. This study was supported by the Plant Molecular Biology Center at Northern Illinois University.

Literature Cited

Ayres NM, AM McClung, PD Larkin, HFJ Bligh, CA Jones, WD Park 1997 Microsatellites and single-nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94:773–781.

Bell CJ, JR Ecker 1994 Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics 19:137–144.

Bennett MD, IJ Leitch 1997 Nuclear DNA amounts in angiosperms: 583 new estimates. Ann Bot 80:169–196.

Bligh HFJ, RI Till, CA Jones 1995 A microsatellite sequence closely linked to the *Waxy* gene of *Oryza sativa*. Euphytica 86:83–85.

Devitt ML, JP Stafstrom 1995 Cell cycle regulation during growth-dormancy cycles in pea axillary buds. Plant Mol Biol 29: 255–265.

Dickey LF, ME Petracek, TT Nguyen, ER Hansen, WF Thompson 1998 Light regulation of *Fed-1* mRNA requires an element in the 5′ untranslated region and correlates with differential polyribosome association. Plant Cell 10:475–484.

Djian P 1998 Evolution of simple repeats in DNA and their relation to human disease. Cell 94:155–160.

Ellis THN, SJ Poyser, MR Knox, AV Vershinin, MJ Ambrose 1998 Polymorphism of insertion sites of Ty1-copia class retrotransposons and its use for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19.

Hoey BK, KR Crowe, VM Jones, NO Polans 1996 A phylogenetic analysis of *Pisum* based on morphological characters, and allozyme and RAPD markers. Theor Appl Genet 92:92–100.

Huang Q, A Beharav, Y Li, V Kirzhner, E Nevo 2002 Mosaic microecological differential stress causes adaptive microsatellite divergence in wild barley, *Hordeum spontaneum*, at Neve Yaar, Israel. Genome 45:1216–1229.

Hüttel B, P Winter, K Weising, W Choumane, F Weigand, G Kahl 1999 Sequence-tagged microsatellite markers for chickpea (*Cicer arietinum* L.). Genome 42:210–217.

Innan H, R Terauchi, NT Miyashita 1997 Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. Genetics 146:1441–1452.

Kalcheva N, HM Lachman, B Shafit-Zagardo 1999 Survey of CAG repeat polymorphisms in the human MAP-2 gene. Psychiatr Genet 9:43–46.

Knox MR, THN Ellis 2001 Stability and inheritance of methylation states at *Pst*I sites in *Pisum*. Mol Genet Genom 265:497–507.

Larkin PD, WD Park 1999 Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. Plant Mol Biol 40:719–727.

Li Y, T Fahima, AB Korol, J Peng, MS Röder, V Kirzhner, A Beiles, E Nevo 2000 Microsatellite diversity correlated with ecological-edaphic and genetic factors in three microsites of wild emmer wheat in north Israel. Mol Biol Evol 17:851–862.

Mathews DH, J Sabina, M Zuker, DH Turner 1999 Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288:911–940.

Morgante M, AM Olivieri 1993 PCR-amplified microsatellites as markers in plant genetics. Plant J 3:75–182.

Murray MG, DL Peters, WF Thompson 1981 Ancient repeated sequences in the pea and mung bean genomes and implications for genome evolution. J Mol Evol 17:31–42.

Nagaraju J, M Kathirvel, RR Kumar, EA Siddiq, SE Hasnain 2002 Genetic analysis of traditional and evolved Basmati and non-Basmati rice varieties by using fluorescence-based ISSR-PCR and SSR markers. Proc Natl Acad Sci USA 99:5836–5841.

Neumann P, M Nouzova, J Macas 2001 Molecular and cytogenetic analysis of repetitive DNA in pea (*Pisum sativum* L.). Genome 44: 716–728.

Pearce SR, M Knox, THN Ellis, AJ Flavell, A Kumar 2000 Pea Ty1-*copia* group retrotransposons: transpositional activity and use as markers to study genetic diversity in Pisum. Mol Gen Genet 263: 898–907.

Polhill RM 1981 Papillionoideae. Pages 191–208 *in* RM Polhill, PH Raven, eds. Advances in legume systematics. Pt 1. Royal Botanic Gardens, Kew.

Raca G, EY Siyanova, CT McMurray, SM Mirkin 2000 Expansion of the (CTG)(n) repeat in the 5′-UTR of a reporter gene impedes translation. Nucleic Acids Res 28:3943–3949.

Reddy PS, DE Housman 1997 The complex pathology of trinucleotide repeats. Curr Opin Cell Biol 9:364–372.

Saghai Maroof MA, RM Biyashev, GP Yang, Q Zhang, RW Allard 1994 Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. Proc Natl Acad Sci USA 91:5466–5470.

Sefc KM, F Regner, E Turetschek, J Glössl, H Steinkellner 1999 Identification of microsatellite sequences in *Vitis riparia* and their

application for genotyping of different *Vitis* species. Genome 42: 367–373.

Sherameti I, M Nakamura, YY Yamamoto, T Pfannschmidt, J Obokata, R Oelmüller 2002 Polyribosome loading of spinach mRNAs for photosystem I subunits is controlled by photosynthetic electron transport. Plant J 32:631–639.

Stafstrom JP, BD Ripley, ML Devitt, B Drake 1998 Dormancy-associated gene expression in pea axillary buds: cloning and expression of PsDRM1 and PsDRM2. Planta 205:547–552.

Stafstrom JP, IM Sussex 1992 Expression of a ribosomal protein gene in axillary buds of pea. Plant Physiol 100:1494–1502.

Tautz D 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471.

Toutenhoofd SL, F Garcia, DA Zacharias, RA Wilson, EE Strehler 1998 Minimum CAG repeat in the human calmodulin-1 gene 5′ untranslated region is required for full expression. Biochim Biophys Acta 1398:315–320.

Udupa SM, LD Robertson, F Weigand, M Baum, G Kahl 1999 Allelic variation at $(TAA)_n$ microsatellite loci in a world collection of chickpea (*Cicer arietinum* L.) germplasm. Mol Gen Genet 261: 354–363.

Wang Z, JL Weber, G Zhong, SD Tanksley 1994 Survey of plant short tandem repeats. Theor Appl Genet 88:1–6.

Yu K, SJ Park, V Poysa 1999 Abundance and variation of microsatellite DNA sequences in beans (*Phaseolus* and *Vigna*). Genome 42:27–34.