Significant Expansion of *Vicia pannonica* Genome Size Mediated by Amplification of a Single Type of Giant Retroelement

Pavel Neumann,*† Andrea Koblížková,* Alice Navrátilová,* and Jiří Macas*†

*Institute of Plant Molecular Biology, České Budějovice 37005, Czech Republic
†Department of Horticulture, University of Wisconsin-Madison, Madison, WI 53706, USA

Sequence data from this article have been deposited with the EMBL/GenBank Data libraries under accession no AY936172.
Running head: Impact of Ogre on Genome size

Key words: Vicia, Ogre elements, LTR retrotransposon, genome size, plant genome evolution

1Corresponding author: Jiří Macas

Institute of Plant Molecular Biology
Branišovská 31
České Budějovice
CZ-37005, Czech Republic

e-mail: macas@umbr.cas.cz
phone: +420 387775516
fax: +420 385310356
ABSTRACT

Amplification and eventual elimination of dispersed repeats, especially those of the retroelement origin, account for most of the profound size variability observed between plant genomes. In most higher plants investigated so far, differential accumulation of various families of elements contribute to these differences. Here we report the identification of giant Ty3/gypsy-like retrotransposons from the legume plant *Vicia pannonica*, which alone make up about 38% percent of this species genome. These retrotransposons have structural features of the Ogre elements previously identified in the genomes of pea and Medicago. These features include extreme size (25 kb), presence of an extra ORF upstream of gag-pol region, and a putative intron dividing the prot and rt coding sequences. The Ogre elements are evenly dispersed on *V. pannonica* chromosomes except for terminal regions containing satellite repeats, their individual copies show extraordinary sequence similarity, and at least part of them are transcriptionally active which suggests their recent amplification. Similar elements were also detected in several other *Vicia* species but in most cases in significantly lower numbers. However, there was no obvious correlation of the abundance of Ogre sequences with the genome size of these species.
INTRODUCTION

Nuclear genomes of higher plants differ considerably in their size, ranging from 0.1 pg (98 Mbp) in *Fragaria viridis* to 89.5 pg (87,686 Mbp) in *Fritillaria davisii* (Bennett and Leitch 2004). Even closely related species belonging to the same genus can display five-to ten-fold differences in their haploid genome size, as it is for example in Phalenopsis, Scilla or Vicia (Bennett and Leitch 2005). First investigations of this phenomenon using DNA reassociation kinetics (Chooi 1971; Flavell *et al.* 1974) revealed that genome size variation is mainly caused by differences in the proportion of repetitive DNA sequences. This was later confirmed by finding many families of repetitive sequences from a number of species. Among these, satellite repeats and retroelements have the most significant impact on genome size. Satellite repeats are organized as long arrays of tandemly repeated units (monomers). Although the monomer sequences are usually only tens to hundreds of nucleotides long (Macas *et al.* 2002), they can be amplified up to millions of copies (Kato *et al.* 1984; Ingham *et al.* 1993; Iriune *et al.* 1995; Macas *et al.* 2000), making up to 20% of the genome (Ingham *et al.* 1993). However, in most plant species investigated so far, the majority of repetitive DNA is composed of various families of retroelements (reviewed in Kumar and Bennetzen 1999; Feschotte *et al.* 2002). This high proportion of retroelements within plant genomes is a consequence of their replicative (copy-and-paste) mode of transposition (retrotransposition), which generates a new copy of the element each time it is retrotransposed. Although the retroelements do not attain as high copy numbers as the satellite repeats, their impact on the genome size is more pronounced due to their considerable length, ranging from a few up to 14 kb (Hirochika *et al.* 1992; Martinez-Izquierdo *et al.* 1997; Neumann *et al.* 2005). The recent discovery of
a new group of giant retrotransposons, named Ogre elements, showed that the upper length limit of retrotransposons could be even longer. These Ty3/gypsy-like elements identified in pea (*Pisum sativum*) and *Medicago truncatula* are up to 22 kb long, and they occur at about 10,000 copies in the pea genome, corresponding to at least 5% of its nuclear DNA (*Neumann et al.* 2003).

The genus Vicia (Fabaceae) includes over 160 species differing considerably in their haploid nuclear DNA content (1.9-14.4 pg, corresponding to 1,862 – 14,112 Mbp) (*Bennett* and *Leitch* 2005). Several studies revealed that there is a number of differentially amplified repeats of the retroelement origin that significantly contribute to these differences (*Pearce et al.* 1996; *Kumar et al.* 1997; *Nouzová et al.* 2001; *Hill et al.* 2005). However, these studies described only partial retroelement sequences, which do not allow precise evaluation of the contribution that specific element families give to the Vicia genomes’ evolution. In this work, we show that many of these partial sequences belong to a retrotransposon family closely related to the giant Ogre elements previously identified in pea and Medicago. We describe full-length Ogre-like elements isolated from the genome of *V. pannonica*, and present data suggesting that significant expansion of the genome size in some Vicia species was caused by recent amplification of these elements.
MATERIALS AND METHODS

Plant material: Seeds of *Vicia pannonica* (cv. Dětěnická panonská) and other species used in this study were obtained from the seed bank of the Institute of Plant Genetics and Crop Plant Research, Gatersleben (Germany), Agritec Šumperk (Czech Republic), and Plant Breeding Station, Boršov (Czech Republic). Total genomic DNA was extracted from leaves as described by DELLAPORTA *et al.* (1983). All DNA concentration measurements were done using the PicoGreen dye (Molecular Probes) according to manufacturer's recommendations. The measurements were performed in microwell plates and PicoGreen fluorescence was evaluated using fluoroimager (Typhoon 9410, Amersham). The whole set of genomic DNAs and control fragments used for dot-blotting was measured simultaneously, using the same DNA concentration standards.

Cloning procedures and sequence analysis: Restriction fragments appearing as bands on ethidium bromide-stained agarose gels of *V. pannonica* DNA digested with *Bgl*II or *Kpn*I were cut out of the gel, purified, and cloned into pBluescript II SK+ (Stratagene) digested with the corresponding enzyme. The fragments were sequenced using the dideoxy-mediated chain-termination method (SANGER *et al.* 1977).

A cosmid library was prepared by partial digestion of high molecular weight genomic DNA of *V. pannonica* with *Mbo*I, followed by its dephosphorylation and cloning into *Bam*HI-digested vector SuperCos1 (Stratagene). The library was screened with the cloned *Bgl*II and *Kpn*I fragments as probes using AlkPhos Direct Hybridization and Detection Kit (Amersham). A clone selected for sequencing (VP-cosC6) was subcloned and the sequencing templates from individual subclones were prepared using
GeneJumper Primer Insertion Kit (Invitrogen). Sequence assembly and basic analysis was done with Staden Package software (Staden 1996). As the cosmid clone contained several highly similar elements, the sequence assembly of the whole clone was verified by its restriction analysis and by PCR with primers specific for its individual parts. Computer analysis of the resulting sequence was performed using the Dotter program (Sonnhammer and Durbin 1995), program tools implemented at the Biology WorkBench website (http://workbench.sdsc.edu/), and EMBOSS (Rice et al. 2000). Multiple sequence comparisons were done using Clustal W (Thompson et al. 1994). BLAST and FASTA (Pearson and Lipman 1988; Altschul et al. 1997) were employed for homology searches and the RPS-Blast (Marchler-Bauer et al. 2003) was used to search for conserved protein domains. Phylogenetic analyses was done using Clustal W and a phylogenetic tree was reconstructed using Phylojava client/server tool. Sequences of reverse transcriptase domains were mostly taken from the alignment ALIGN_000602 (Vicient et al. 2001). Sequences of tRNAs used for identification of pbs (primer binding site) were obtained from the Arabidopsis thaliana tRNA database (Lowe and Eddy 1997). Splice site analysis was performed at NetGene2 server (Hegade et al. 1996).

**Copy-number estimation:** To estimate the copy number of Ogre elements in the genomes of selected species, serial dilutions of their genomic DNAs were quantitatively dot-blotted on Hybond N+ membrane (Amersham) together with fragments of the Ogre element as hybridization standards. The quantity of spotted genomic DNA corresponding to 50 – 10^5 copies of the haploid genomes (1C) were compared to 5x10^6 - 1x10^10 copies of the hybridization standards. Dot-bLOTS were hybridized with the probes specific for different regions of the Ogre elements (LTR, ORF1, 2, and 3) prepared by PCR using VP-
cosC6 DNA as a template and employing the following primers: VP1 5’ AAC TTT TAG TCA TTT ACT TTC AAT AAA CA 3’ and MT-pbs 5’ TCC CCA GTG AAG TCG CCA 3’; VP17 5’ TGG GAA GAA GAA ACA CCA AG 3’ and VP18 5’ CAT CTT CAT TTG ACG AGC AA 3’; VP19 5’ AAC GAG CTT CGT GGT ACA AT 3’ and VP20 5’ CTC GAG GAT TGT TGT GAC AG 3’; VP23 5’ CGA AGA GGA TGA AGA AGA GG 3’ and VP24 5’ TTT CTT GAC TGC ATC AGC AT 3’. Probe positions in Ogre-VP1 element are shown in the Figure 2A. Probe labeling and hybridization were done using the AlkPhos kit following the manufacturer's instructions. Hybridization and washing temperatures were 65°. Probe detection was performed using CDP-Star substrate (Amersham) and the signals were captured on X-ray film. The signals were also captured and quantified using Typhoon 9410 scanner and ImageQuant TL software (Amersham).

Alternatively, estimation of Ogre copy numbers in *V. pannonica* genome was based on the number of positive clones observed after screening a short-insert shotgun genomic library with the probes described above. The shotgun library was prepared from *V. pannonica* DNA subjected to sonication, mung bean nuclease treatment and size fractionation on an agarose gel. The 600-800 bp fragments were purified from the gel, treated with polynucleotide kinase in the presence of ATP, and cloned into dephosphorylated *Sma*I-cut plasmid vector (pBluescript II SK+). Copy number (CN) was calculated for individual probes using the formula CN = GS x PG / Lp, where GS is the genome size of *V. pannonica* (6.51 x 10⁹ bp), PG is the proportion of the probe sequence in the genome, and Lp is the probe length. The PG value was defined as PG = Lh / Lt, where Lh is the length of hybridizing sequences within the library and Lt is the sum of insert lengths of all screened clones. As it was supposed that many of positive clones did not hybridize over their whole length, the Lh value could not be calculated by simply
multiplying the number of positive clones with the average insert length. Instead, the Lh value was calculated considering the theoretical frequency of clones hybridizing over their whole length (F1), the theoretical frequency of clones hybridizing with only part of their sequences (F2), the number of positively hybridizing clones (Np), the average insert size (Li = 700 bp), the minimum length of sequence capable of efficient hybridization (Lmin = 100 bp), and the average length of partially hybridizing sequences (Lavg = (Li – Lmin – 1) / 2), using the formula \( Lh = Np \times F1 \times Li + Np \times F2 \times Lavg \). The F1 and F2 values were calculated using the formulas \( F1 = (Lp - Li) / ((Lp - Li) + 2 \times (Li - Lmin - 1)) \) and \( F2 = 2 \times (Li - Lmin - 1) / ((Lp - Li) + 2 \times (Li - Lmin - 1)) \).

**RNA isolation and RT-PCR:** The tissues used for RNA isolation (leaf, root, flower) were harvested from plants cultivated in pots under a 15 h light / 9 h dark photoperiod at 22 and 18\(^\circ\), respectively. Total RNA was isolated using the Total RNA Isolation Kit (Ambion). All RNA samples were treated with RNase-free DNase (Invitrogen) to remove any contaminant DNA. Reverse transcription was carried out using the SuperScript II Reverse Transcription Kit (Invitrogen) by a random priming method according to manufacturer's recommendations, using 0.5 \( \mu \)g of template RNA. The RT-PCR reaction mix (25 \( \mu \)l) consisted of 1 x PCR buffer, 0.2 mM dNTPs, 0.2 \( \mu \)M primers, 1.5 mM MgCl\(_2\), 1 U of Platinum *Taq* polymerase (Invitrogen) and 2.5 ng of reverse-transcribed RNA or an equal amount of reverse transcriptase-untreated RNA as a negative control. The following primers were used in the RT-PCR experiments: VP17 and VP18; VP25 5’ ACG TTC TCT TTC ATC GAT GC 3’ and MT3 5’ CGG TAG TCA ACA CAC ATT CTG AC 3’. The reaction profile included 35 cycles of 30 sec at 94\(^\circ\), 50 sec at 55\(^\circ\) and 1-3 min at 72\(^\circ\); preceded by initial denaturation (3 min at 94\(^\circ\)) and followed by final extension
step (10 min at 72°C). Reaction products were resolved on agarose gel electrophoresis.

**Fluorescence in situ hybridization:** FISH was performed on isolated chromosomes prepared as described by Gualberti et al. (1996) and centrifuged onto slides using a Hettich centrifuge equipped with cytospin chambers. The Ogre probe was derived from the longest EcoRV restriction fragment of the clone VP-cosC6 containing LTR, 5′UTR, ORF1 and ORF2 regions. The probe was labeled with biotin-16-dUTP (Boehringer Mannheim) using random priming and detected using streptavidin-Alexa Fluor-568 (Molecular Probes) as described by Leitch et al. (1994). Treatment of slides before hybridization, composition of hybridization mix and hybridization conditions of the Ogre probe were as described by Neumann et al. (2001). Following hybridization of the Ogre probe, the slides were dehydrated using the ethanol series, air-dried and used for the second round of hybridization with fluorescein-labeled oligonucleotide probe (5′ AAG ATT RTC TTG TGY TAT AST ACA TAA AAK TCA CGA AGT 3′) specific for satellite repeats VicTR-A (Macas et al. 2000), which produce labeling patterns allowing discrimination of all chromosome types within V. pannonica karyotype (Navrátilová et al. 2003). Hybridization of the VicTR-A probe (0.5 ng/µl) was performed at 37°C for 16 hours in the hybridization mix consisting of 2x SSC, 100 ng/µl sheared calf thymus DNA, and 0.125% SDS, and post-hybridization washes were done in 2x SSC at 42°C for 10 min. Chromosomes were counterstained with DAPI (4′,6′-diamidino-2-phenylindole) and examined using a Nikon Eclipse 600 microscope equipped with appropriate filter sets. The images were captured with a CCD camera and analyzed using Lucia software (Laboratory Imaging).
RESULTS

In our preliminary experiments aimed at comparative analysis of Vicia genomes, we noticed that *V. pannonica* produces characteristic restriction patterns when its DNA is digested with restriction endonucleases and resolved on agarose gels. Contrary to the majority of species used for the experiments, which produced smears with a few distinct bands, these patterns consisted of a series of well distinguished bands in most enzymes tested (Figure 1A). This observation suggested that there is a prominent and conserved DNA repeat making up a large portion of the *V. pannonica* genome. In order to isolate and characterize this repeat, we cloned four of the distinct bands from the DNA digested with *Bgl*II (850 and 1250 bp) or *Kpn*I (560 and 1000 bp). Sequence analysis of the cloned fragments revealed their similarity to different parts of the giant retroelement Ogre previously described in pea (Neumann et al. 2003). To isolate corresponding full-length element from *V. pannonica*, a cosmid library was constructed and screened with the cloned fragments as probes. Thirty randomly picked clones hybridizing to all four probes were subjected to restriction analysis in order to check for the presence of fragments corresponding to the conserved genomic bands. These fragments were found in most of the clones, thirteen of which (43%) produced bands corresponding to all four fragments used as probes. One such clone (VP-cosC6) was selected for complete sequencing and further analysis.

**Sequence characterization and transcriptional activity of Ogre elements in *V. pannonica*:** The insert cloned in VP-cosC6 (GenBank accession number AY936172) was over 42 kb in length and entirely composed of Ogre sequences belonging to three
different elements (Figure 2). There was one complete element, designated Ogre-VP1, which contained intact LTRs surrounding internal coding regions and was flanked by a 5-bp target site duplication (5’ ATGCC / ATGCC 3’). This element was inserted into another Ogre element (Ogre-VP2), whose sequence was truncated at its left LTR due to cloning and at its right LTR due to insertion of the third element, Ogre-VP3. Most of the Ogre-VP3 sequence was lost through cloning except for a part of its left LTR. All three elements shared high mutual similarities. The overall similarity between Ogre-VP1 and Ogre-VP2 was 92.3%. The Ogre-VP3 LTR shared 92.1% similarity with the LTR sequence of Ogre-VP1 and 81.5% similarity with LTR of Ogre-VP2. Similarity of these elements to the partial sequences cloned as conserved BglII and KpnI fragments was 86-99%.

The only full-length element, Ogre-VP1, was 25,049 bp long, which makes it the largest plant retroelement described so far. This extreme size was mainly due to exceptionally long LTRs, each of which spans 6,438 bp. Sequence similarity between the LTRs was 99.9% (only 6 nucleotide substitutions over their entire length), which indicates that Ogre-VP1 represents a very recent insertion. The overall arrangement of Ogre-VP1 coding and structural regions is identical to that of the Ogre elements identified in the *Pisum sativum* and *Medicago truncatula* genomes (Neumann et al. 2003) and further confirms its assignment into this group of Ty3/gypsy-like retrotransposons. Close relationships of Ogre sequences from all three species were also apparent from a phylogenetic analysis of the reverse transcriptase protein domains (supplementary Figure S1 at http://www.genetics.org/supplemental/). The coding region of Ogre-VP1 is divided into three reading frames, (ORF1-3, Figure 2A) separated by short regions containing several stop codons in all three frames. These ORFs could be directly translated into
protein sequences, with the exception of ORF3, containing a +1 frameshift due to an insertion of one nucleotide, which had to be removed to allow its conceptual translation.

Similarly to the pea elements, ORF1 of Ogre-VP1 encoded a protein with unknown function. This ORF was 1527 bp in length and was separated from ORF2 by a 237 bp long non-coding region containing stop codons. The partial Ogre-VP2 sequence contained an ORF1 of the same length and with 97% sequence similarity to ORF1 in Ogre-VP1. All of the protein domains typical for plant retroelements could be recognized within the following two ORFs (ORF2 and ORF3). Gag and protease (prot) domains were encoded by ORF2, whereas reverse transcriptase (rt), ribonuclease H (rh) and integrase (int) were encoded by ORF3 (Table 1). These ORFs were separated by a region of 317 bp, which contained multiple stop codons. The fact that position of this region corresponded to that of an intron within the pea and Medicago truncatula Ogre sequences (Neumann et al. 2003, and our unpublished data) in addition to further computer analysis using NetGene2 server (Hebsgaard et al. 1996) strongly suggested that this region could also represent an intron sequence. Moreover, the eventual removal of this predicted intron sequence by splicing would result in joining of ORF2 (+2 frame; 4403 bp) with ORF3 (+1 frame; 3464 bp) into one frame encoding a polyprotein of 2622 amino acids.

Transcriptional activity of V. pannonica Ogre elements was tested using RT-PCR with total RNA isolated from leaves, roots and flowers. Primers were designed to amplify ORF1 sequence as this region is specific for Ogre elements and it is absent in all other types of retrotransposons described so far. The products of expected length were detected in all three organs tested (Figure 3A). In order to test the splicing of the predicted intron between ORF2 and ORF3, primers directed towards this region were also used. Although the corresponding RNA was again detected in all samples, the size of the amplified
fragments did not correspond to the spliced sequence (Figure 3B). Cloning and sequence analysis of these fragments revealed 92 – 97% similarity to the corresponding region in Ogre-VP1. Comparison of the RT-PCR-amplified sequences with genomic sequences confirmed that none of them was spliced.

**Abundance of Ogre-like sequences in Vicia and other legume species:** The copy number of Ogre elements in the *V. pannonica* genome was measured using two different approaches. The first estimate was based on quantitative dot-blot hybridization of serial dilutions of genomic DNA and control fragments (cloned Ogre sequences) with probes derived from four regions of the element (see Figure 2A for the probe positions). Resulting signals corresponded to 1-5x10^5 copies per haploid genome (1C) for the LTR probe and to 1-2x10^5 copies/1C for each of the remaining probes (ORF1, ORF2 and ORF3). The second method employed hybridization of probes for LTR, ORF1 or ORF2, to a shotgun, short-insert library of *V. pannonica* genomic DNA. Out of 6,530 clones screened, 409, 169 or 237 clones, respectively, displayed positive hybridization signal. Considering the length of individual probes (1615, 1518, and 1999 bp) and using formulas described in the materials and methods, the estimated copy number per 1C was 2x10^5 for LTR, 8x10^4 for ORF1 and 9x10^4 for ORF2. Thus, the copy numbers calculations based on these two principally different methods were in good accordance and a copy number of 1x10^5 copies/1C can be considered a reliable estimation. Taking into account the genome size of *Vicia pannonica* (6.75 pg, Raina and Rees 1983) and assuming that all element copies are full-length, the copy number of 10^5/1C corresponds to 38% of the genome being made up by Ogre sequences.

In order to test if such a high amplification of Ogre elements resulted in their
accumulation in specific genomic regions or if it was accompanied by an overall increase of the element copies throughout the entire genome, we performed detection of the Ogre repeats on mitotic chromosomes using in situ hybridization. These experiments showed that the elements were dispersed over the entire genome, spanning all of the chromosomes (Figure 4). The only exception to this homogenous chromosome labeling were the subtelomeric regions of the short chromosome arms which produced weaker signals due to the presence of the highly amplified satellite repeat VicTR-A (MACAS et al. 2000).

The quantitative dot-blot hybridization was also used to estimate the abundance of Ogre-like sequences in twelve Vicia species and in several other legumes (Table 2). Two of these species, V. melanops and V. hybrida, were found to contain highly amplified Ogre elements, which were estimated to reach up to $1 \times 10^5$ and $1-5 \times 10^4$ copies/1C, respectively. These findings were in agreement with the results of Southern blot experiments in which only these two species showed prominent signals in addition to V. pannonica (Figure 5). Other Vicia species produced hybridization signals corresponding to lower numbers of Ogre sequences; however, the numbers estimated for individual species varied depending on the probe used (Table 2). This observation most likely reflects sequence divergence of the corresponding regions (e.g. ORF1) among species, or eventual cross-hybridization of probes derived from regions conserved among various groups of retroelements (ORF3). This is even more evident for the estimates made for Pisum sativum, which were considerably lower than those made using Pisum-derived Ogre probes (NEUMANN et al. 2003 and Table 2). The species from other genera, including Vigna unquiculata, Cicer arietinum, Glycine max, Lotus angustifolius, Phaseolus vulgaris, Medicago truncatula, produced no or very weak hybridization, not exceeding
signals corresponding to less than 100-500 copies per 1 C (unpublished data).

Besides previously described Ogre sequences (Neumann et al. 2003), a number of partial clones of repetitive elements from *V. melanops*, *V. sativa* and *V. narbonensis* were identified as most similar to Ogre-VP elements in BLAST and FASTA homology searches (Table 3). The high sequence similarities strongly suggest that these clones correspond to various regions of Ogre-like elements, including ORF1.
DISCUSSION

A crucial role of the amplification and eventual elimination of retroelement sequences in plant genome evolution has become evident through the accumulation of sequence data and subsequent comparative analysis of species with different genome sizes. For example, a comparison of the Adh1-containing orthologous regions in maize and sorghum demonstrated that while more than 60% of this region in maize is composed of LTR retrotransposons, no retrotransposons were detected in the orthologous region in sorghum (SanMiguel et al. 1996; SanMiguel and Bennetzen 1998). These LTR retrotransposons constitute 49 – 78% of the maize genome suggesting that the three-fold greater genome size of maize compared to sorghum was a result of retrotransposon amplification (SanMiguel and Bennetzen 1998). Similar role of retroelements in genome size evolution has also been revealed in other species (Shirasu et al. 2000; Wicker et al. 2001; Wicker et al. 2005). In general, all high copy number retrotransposons were identified in species with large genomes (Pearce et al. 1996; Pearce et al. 1997; Meyers et al. 2001; Muniz et al. 2001; Kentner et al. 2003) while species with small genomes contain these elements at relatively low copy numbers. The latter is the case of well characterized genomes of Arabidopsis thaliana (130 Mb) and Oryza sativa (430 Mb) in which retrotransposons make up only 4% and 22% of the genome (Ma et al. 2004; Peterson-Burch et al. 2004; Zhang and Wessler 2004). All plant genomes characterized so far are composed of many diverse families of LTR retrotransposons with different degrees of amplification. The copy number of elements belonging to one family can vary greatly from a few to about $10^5$ copies per genome (Kumar and Bennetzen 1999; Feschotte et al. 2002). The most abundant plant LTR retrotransposons, such as PREM-2,
Opie or Huck-2 in Zea mays and IRRE in Iris, can make up to 10% of the genome (SANMIGUEL and BENNETZEN 1998; MEYERS et al. 2001; KENTNER et al. 2003; CHANRET et al. 2005). Thus, concurrent amplification of elements belonging to several families can account for considerable increases in genome size as demonstrated for Zea species (SANMIGUEL and BENNETZEN 1998; MEYERS et al. 2001). Our results described here demonstrate that even the amplification of a single retroelement family can increase the genome size by more than 50% (based on the calculation that this element makes up 38% of the V. pannonica genome). Although the Ogre sequences were detected in all Vicia species tested, their abundance differed by several orders of magnitude (Table 2). Whereas the copy numbers in some species could be underestimated due to sequence divergence of the elements, this great variability in copy numbers was evident even among closely related species each belonging to the taxonomic section Hypechusa. Among the four species tested, the Ogre sequences were about 10-fold less abundant in V. lutea than in V. hybrida, and 100-fold less abundant than in V. pannonica and V. melanops. Interestingly, the genome size of V. lutea is similar to those of the other three species, suggesting that other repetitive elements may have been amplified there instead of Ogre. The amplification of other elements probably accompanied the expansion of the genome size in V. melanops, as the copy number of Ogre family in this species is the same as in V. pannonica, but its genome is larger (Table 2). Therefore, although it is evident from the differences in Ogre copy numbers in individual species that these elements have played a crucial role in the increase of genome size for at least some Vicia species, there is no simple correlation between abundance of Ogre sequences and the genome size in this genus.

The differences in copy numbers observed among Vicia species could be in part
explained by a loss of Ogre sequences in some genomes due to unequal homologous recombination and illegitimate recombination, which were shown to be responsible for deletions of retrotransposon sequences from genomes in several plant species (Shirasu et al. 2000; Devos et al. 2002; Vitte and Panaud 2003; Wicker et al. 2003; Ma et al. 2004).

An indicator of unequal recombination is a presence of solo LTRs, which remain in a genome after a recombination event. The copy number estimated for LTR sequences was in several species considerably higher compared to inside regions (e.g. Vicia lutea and V. hybrida, Table 2), although the LTR sequences are believed to be less conserved and thus less likely to hybridize with a probe from another species than the internal coding sequences. Thus, it is possible that in these species the reduction of the genome size through the elimination of Ogre sequences took place, leaving an excess of solo LTRs.

Molecular mechanisms facilitating the eventual regulation of retroelement amplification in plant genomes are only poorly understood. Mechanisms, which are likely to play a role in the suppression of mobile elements include transcriptional silencing through DNA methylation (Hirochika et al. 2000; Wright and Voytas 2002; Liu et al. 2004), chromatin modifications (Jackson et al. 2002), and posttranscriptional silencing by RNA interference mechanism (Timmonts 2002). The high amplification of Ogre in a limited number of species and its occurrence in only moderate copy numbers in others can be explained either by an insufficient function of mechanisms suppressing transposition of retroelements, or by escaping of Ogre elements from such mechanisms. Since virtually all prominent bands visible on digested genomic DNA of V. pannonica hybridized to Ogre, it seems that only this family of retrotransposons was amplified to a high copy numbers, whereas other retroelements remained suppressed. Considering the high sequence similarity among Ogre sequences isolated from V. pannonica, the high
level of conservation in restriction sites, and the recent insertion of Ogre-VP1 element (inferred from the similarity of its LTRs), it is likely that the burst of amplification of the Ogre family happened quite recently. As these elements are still transcribed in *V. pannonica*, the amplification of the Ogre family in this species may not be over yet. In this respect, it is interesting to mention that the transcription of Ogre elements was detected also in *Pisum sativum* (Neumann et al. 2003) containing one order of magnitude fewer copies compared to *V. pannonica*, and in *Medicago truncatula* having small genome (0.48 pg or 466 Mbp; Arumuganathan and Earle 1991) which, consequently, can be occupied by relatively few Ogre elements. This could imply that transcription itself is not sufficient for transpositional activity and that post-transcriptional regulation plays an important role in the suppression of transposition.

From the results described here, as well as from our previous findings (Neumann et al. 2003), it is evident that Ogre elements represent a group of retrotransposons capable of inducing significant changes in the genome size of some plant species. Up to now, full-length Ogre elements have been detected only in a relatively narrow range of legume taxa including the genera of *Pisum*, *Medicago*, and *Vicia*. However, except for the hybridization-based detection described here no attempts were performed to identify related elements in other plants. Moreover, no or very weak hybridization signals obtained in species from more distant genera (*Vigna*, *Cicer*, *Glycine*, *Lotus*, and *Phaseolus*) does not necessarily indicate the absence of Ogre elements but could also be caused by their further sequence divergence. Thus, the alternative approaches based on sequence analysis of regions typical for Ogre family including mainly ORF1 and pbs, should be employed to find Ogre-like elements in other species. Future experiments should also be directed to structural and functional analysis of the extremely long Ogre
LTRs harboring regulatory regions controlling the element replication cycle, which could provide clues for explaining high amplification rate of Ogre elements in some species.

**Acknowledgment:** We thank Ms. H. Štěpančíková for excellent technical assistance, Dr. M. Nouzová for help with DNA sequencing, and Ms. S. M. Rafelski and Lara Colton for assistance in preparation of the manuscript. This work was supported by grants GACR521/00/0655 and GACR521/02/P007 from the Grant Agency of the Czech Republic, and AVOZ50510513 from the Academy of Sciences of the Czech Republic.
LITERATURE CITED


CHANTRET, N., J. SALSE, F. SABOT, S. RAHMAN, A. BELLEC et al., 2005 Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (Triticum and aegilops). Plant Cell 17: 1033-1045.


HIROCHIKA, H., H. OKAMOTO and T. KAKUTANI, 2000 Silencing of retrotransposons in


Liu, Z. L., F. P. Han, M. Tan, X. H. Shan, Y. Z. Dong et al., 2004 Activation of a rice endogenous retrotransposon Tos17 in tissue culture is accompanied by cytosine


PEARCE, S. R., G. HARRISON, D. T. LI, J. S. HESLOPHARRISON, A. KUMAR et al., 1996 The Ty1-


SANMIGUEL, P., and J. L. BENNETZEN, 1998 Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. Ann. Bot. 82: 37-44.


VICIENT, C. M., R. KALENDAR and A. H. SCHULMAN, 2001 Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Research 11: 2041-2049.


WICKER, T., W. ZIMMERMANN, D. PEROVIC, A. H. PATTERSON, M. GANAL et al., 2005 A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. Plant J.
41: 184-194.


Table 1. Conserved domains found in the putative polyprotein encoded by Ogre-VP1

<table>
<thead>
<tr>
<th>CD</th>
<th>E-value</th>
<th>Position</th>
<th>Active site motif</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>gag (pfam03732) prot</td>
<td>3e-09</td>
<td>264-350</td>
<td></td>
<td>The domain is followed by putative zinc finger motif</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CX$_2$HX$_2$HX$_4$C (513-527)$_v$</td>
</tr>
<tr>
<td>rt (pfam00078)</td>
<td>7e-24</td>
<td>1624-1793</td>
<td>YVDD$_v$</td>
<td>The domain was identified based on similarity to pea Ogre element (acc.no. AY299398)</td>
</tr>
<tr>
<td>rh (pfam00075)</td>
<td>1e-08</td>
<td>2053-2173</td>
<td>DEDD$_v$</td>
<td></td>
</tr>
<tr>
<td>int (pfam00665)</td>
<td>2e-21</td>
<td>2334-2488</td>
<td>DDE$_v$</td>
<td>The domain is preceded by zinc binding motif HHCC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(2280-2314)$_v$</td>
</tr>
</tbody>
</table>

$^a$ Domains typical for LTR retrotransposon polyproteins were identified by searching conserved domain (CD) database using RPS-Blast (Marchler-Bauer et al. 2003) and by similarity to the putative protein encoded by the pea Ogre element identified in the clone Ps-cos16 (acc.no. AY299398).

$^b$ The putative polyprotein sequence was deduced from ORF2 and ORF3, which were joined after the removal of predicted intron.

$^c$ (Skalka 1989)

$^d$ (Xiong and Eickbush 1990; Ding et al. 1998)

$^e$ (Davies et al. 1991; Malik and Eickbush 2001)

$^f$ (Burd and Dreyfuss 1994)

$^g$ (Asante-Apiah and Skalka 1997; Maigan et al. 1998)
Table 2. Estimated copy numbers of Ogre sequences per haploid genomes (1C) of selected Vicia species and Pisum sativum

<table>
<thead>
<tr>
<th>Section</th>
<th>Species</th>
<th>1C</th>
<th>1C</th>
<th>Dot-blot probes</th>
<th>Other estimates</th>
<th>Proportion in the genome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>[Mbp]</td>
<td>[pg]</td>
<td>LTR</td>
<td>ORF1</td>
<td>ORF2</td>
</tr>
<tr>
<td>Hypechusa</td>
<td>V. pannonica</td>
<td>6.615</td>
<td>6.75</td>
<td>1-5x10^5</td>
<td>1x10^3</td>
<td>1x10^5</td>
</tr>
<tr>
<td></td>
<td>V. hybrida</td>
<td>6.640</td>
<td>6.78</td>
<td>1x10^4</td>
<td>1x10^4</td>
<td>1-5x10^4</td>
</tr>
<tr>
<td></td>
<td>V. lutea</td>
<td>7.252</td>
<td>7.40</td>
<td>0.5-1x10^4</td>
<td>0.5-1x10^3</td>
<td>5x10^3</td>
</tr>
<tr>
<td></td>
<td>V. melanos</td>
<td>9.800</td>
<td>10.00</td>
<td>5x10^5</td>
<td>1x10^3</td>
<td>1x10^5</td>
</tr>
<tr>
<td>Cracca</td>
<td>V. villosa</td>
<td>2.230</td>
<td>2.28</td>
<td>5x10^3</td>
<td>0.5-1x10^3</td>
<td>1x10^3</td>
</tr>
<tr>
<td>Attosa</td>
<td>V. sepium</td>
<td>4.582</td>
<td>4.68</td>
<td>1x10^3</td>
<td>1x10^3</td>
<td>5x10^3</td>
</tr>
<tr>
<td>Vicia</td>
<td>V. sativa</td>
<td>2.205</td>
<td>2.25</td>
<td>5x10^3</td>
<td>1-5x10^3</td>
<td>1-5x10^3</td>
</tr>
<tr>
<td></td>
<td>V. grandiflora</td>
<td>3.283</td>
<td>3.35</td>
<td>0.5-1x10^4</td>
<td>1x10^3</td>
<td>5x10^3</td>
</tr>
<tr>
<td>Wiggersia</td>
<td>V. lathyroides</td>
<td>2.573</td>
<td>2.63</td>
<td>1x10^2</td>
<td>0-50</td>
<td>0.5-1x10^2</td>
</tr>
<tr>
<td>Narbonensis</td>
<td>V. narbonensis</td>
<td>7.130</td>
<td>7.28</td>
<td>1-5x10^3</td>
<td>1x10^3</td>
<td>1-5x10^3</td>
</tr>
<tr>
<td>Faba</td>
<td>V. faba</td>
<td>13.25</td>
<td>13.50</td>
<td>1x10^3</td>
<td>0.5-1x10^3</td>
<td>1-5x10^3</td>
</tr>
<tr>
<td>Peregrina</td>
<td>V. peregrina</td>
<td>9.286</td>
<td>9.48</td>
<td>1-5x10^3</td>
<td>1-5x10^3</td>
<td>1x10^3</td>
</tr>
<tr>
<td></td>
<td>V. michauxii</td>
<td>8.134</td>
<td>8.30</td>
<td>1x10^3</td>
<td>1x10^3</td>
<td>1x10^3</td>
</tr>
<tr>
<td>Pisum sativum</td>
<td></td>
<td>4.337</td>
<td>4.43</td>
<td>1-5x10^3</td>
<td>5x10^2</td>
<td>1-5x10^3</td>
</tr>
</tbody>
</table>

a Based on the number of positively hybridizing clones in the V. pannonica library.
b According to NOUZOVA et al. 2001.
c According to NEUMANN et al. 2003.
d Calculations were based on estimated copy numbers of ORF2 probe. Values in parentheses show genome proportions calculated using estimates obtained with other species-specific Ogre probes (as shown in the column “Other estimates”).
Table 3. Ogre-like sequences identified in Vicia species by FASTA sequence similarity searches using Ogre-VP1 as a query

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession number</th>
<th>Similarity [%/length]</th>
<th>E-value</th>
<th>Region</th>
<th>Copy numbers [copies/1C] a</th>
</tr>
</thead>
<tbody>
<tr>
<td>V. melanops</td>
<td>AJ391781</td>
<td>88/666</td>
<td>1.6 x 10^{-70}</td>
<td>LTR</td>
<td>10^5-10^8</td>
</tr>
<tr>
<td></td>
<td>AJ391800</td>
<td>75/982</td>
<td>1.7 x 10^{-86}</td>
<td>LTR</td>
<td>10^4-10^5</td>
</tr>
<tr>
<td></td>
<td>AJ391797</td>
<td>78/493</td>
<td>1.1 x 10^{-37}</td>
<td>ORF1 / spacer</td>
<td>10^4-10^6</td>
</tr>
<tr>
<td></td>
<td>AJ391792</td>
<td>80/479</td>
<td>1.2 x 10^{-94}</td>
<td>ORF2</td>
<td>10^4-10^6</td>
</tr>
<tr>
<td></td>
<td>AJ391793</td>
<td>86/222</td>
<td>2.6 x 10^{-46}</td>
<td>ORF2</td>
<td>10^4-10^6</td>
</tr>
<tr>
<td></td>
<td>AJ391794</td>
<td>90/337</td>
<td>4.5 x 10^{-44}</td>
<td>ORF2</td>
<td>10^4-10^6</td>
</tr>
<tr>
<td></td>
<td>AJ391802</td>
<td>75/211</td>
<td>2.2 x 10^{-30}</td>
<td>ORF2</td>
<td>10^4-10^5</td>
</tr>
<tr>
<td></td>
<td>AJ391799</td>
<td>85/332</td>
<td>6.9 x 10^{-62}</td>
<td>ORF2 / intron</td>
<td>10^4-10^6</td>
</tr>
<tr>
<td></td>
<td>AJ391754, AJ391755,</td>
<td>87/490 - 92/757</td>
<td>2.3 x 10^{-105}</td>
<td>ORF3</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>AJ391759, AJ391760,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AJ391761, AJ391764,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AJ391765</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AJ391774</td>
<td>91/308</td>
<td>7.3 x 10^{-49}</td>
<td>ORF3 / 3' UTR</td>
<td>unknown</td>
</tr>
<tr>
<td>V. sativa</td>
<td>AJ391782</td>
<td>69/781</td>
<td>4.8 x 10^{-30}</td>
<td>LTR</td>
<td>10^4-10^5</td>
</tr>
<tr>
<td></td>
<td>AJ391804</td>
<td>64/118</td>
<td>8.1</td>
<td>5' UTR</td>
<td>10^4</td>
</tr>
<tr>
<td></td>
<td>AJ391758, AJ391762,</td>
<td>68/275 - 87/421</td>
<td>9.1 x 10^{-27}</td>
<td>ORF3</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>AJ391763, AJ391766,</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AJ391767, AJ391769</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AJ391776, AJ391778</td>
<td>82/65, 85/65</td>
<td>6 x 10^{-6}, 7.8 x 10^{-5}</td>
<td>3' UTR</td>
<td>unknown</td>
</tr>
<tr>
<td>V. narbonensis</td>
<td>AJ391777</td>
<td>62/320</td>
<td>1.9 x 10^{-13}</td>
<td>ORF1</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>AJ391756, AJ391757,</td>
<td>70/382 – 68/664</td>
<td>7.3 x 10^{-71}</td>
<td>ORF3</td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td>AJ391768, AJ391770</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AJ391777</td>
<td>76/107</td>
<td>0.0029</td>
<td>3' UTR</td>
<td>unknown</td>
</tr>
</tbody>
</table>

a Copy numbers according to NOUZOVÁ et al. 2001
FIGURE LEGENDS

FIGURE 1.—*Vicia pannonica* Ogre sequences are highly conserved in restriction sites. (A) Genomic DNA of *Vicia pannonica* digested with *TaqI* (T), *RsaI* (R), *EcoRI* (I), *EcoRV* (V), *BglII* (B) and *KpnI* (K). Lambda DNA digested with *PstI* was used as a size marker (M). (B) Southern blot hybridization of the gel shown on panel A using the whole VP-cosC6 insert sequence as a probe.

FIGURE 2.—Ogre elements identified in *Vicia pannonica*. (A) Structure of the full-length element Ogre-VP1. LTRs are depicted as white arrows in black boxes. The open reading frames ORF1, ORF2 and ORF3 are represented by white boxes marked with corresponding numbers. Light gray boxes indicate regions separating individual ORFs and dark gray boxes represent putative 5' and 3' untranslated regions (UTRs). Lines above the scheme show positions of probes used for hybridizations. The horizontal black arrows mark positions and directions (5'->3') of primers used for RT-PCR experiments. Positions of stop codons in three reading frames of the Ogre-VP1 sequence are shown as vertical lines below the scheme. The vertical arrow indicates the frameshift in the ORF3. (B) Structure of the clone VP-cosC6 (GenBank acc.no. AY936172) showing positions and orientation of Ogre-VP1 (black), Ogre-VP2 (dark gray), and Ogre-VP3 (light gray) elements. White arrows represent LTRs.

FIGURE 3.—Transcriptional activity of Ogre retrotransposons in *V. pannonica* analyzed by RT-PCR. The reactions were performed on total RNA isolated from roots, leaves or flowers (RT+). Negative controls with omitted reverse transcriptase step were included in or-
order to check for false positive results (RT-). (A) Amplification of the ORF1 sequence using primers VP17 and VP18. (B) Amplification of the region containing putative intron using primers VP25 and MT3.

**Figure 4.**—FISH on metaphase chromosomes of *V. pannonica*. Two images of each chromosome type are shown. The left image shows DAPI-stained chromosomes (gray) hybridized with the VicTR-A repeat (white signals). The right image shows the same chromosome hybridized with the probe specific for Ogre-VP1.

**Figure 5.**—Agarose gel electrophoresis (A) and corresponding Southern blot (B) of genomic DNAs of selected Vicia species digested with *TaqI* and probed with the whole VP-cosC6 insert sequence.
Figure 2
Figure 3
Figure 4
Figure 5