Theor Appl Genet (2008) 116:859–867
DOI 10.1007/s00122-008-0719-1

ORIGINAL PAPER

A LTR *copia* retrotransposon and *Mutator* transposons interrupt *Pgip* genes in cultivated and wild wheats

Michela Di Giovanni · Alberto Cenci · Michela Janni · Renato D'Ovidio

Received: 18 September 2007 / Accepted: 17 January 2008 / Published online: 27 February 2008

Abstract Polygalacturonase-inhibiting proteins (PGIPs) are leucine-rich repeat (LRR) proteins involved in plant defence. Wheat pgip genes have been isolated from the B (*Tapgip1*) and D (*Tapgip2*) genomes, and now we report the identification of pgip genes from the A genomes of wild and cultivated wheats. By Southern blots and sequence analysis of BAC clones we demonstrated that wheat contains a single copy pgip gene per genome and the one from the A genome, *pgip3*, is inactivated by the insertion of a long terminal repeat *copia* retrotranspon within the fourth LRR. We demonstrated also that this retrotransposon insertion is present in *Triticum urartu* and all the polyploidy wheats assayed, but is absent in *T. monococcum* (*Tmpgip3*), suggesting that this insertion took place after the divergence between *T. monococcum* and *T. urartu*, but before the formation of the polyploid wheats. We identified also two independent insertion events of new Class II transposable elements, *Vacuna*, belonging to the *Mutator* superfamily, that interrupted the *Tdipgip1* gene of *T. turgidum* ssp. *dicoccoides*. The occurrence of these transposons within the coding region of *Tdipgip1* facilitated the mapping of the *Pgip* locus in the pericentric region of the short arm of chromosome group 7. We speculate that the inactivation of pgip genes are tolerated because of redundancy of PGIP activities in the wheat genome.

Communicated by P. Langridge.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-008-0719-1) contains supplementary material, which is available to authorized users.

M. Di Giovanni · M. Janni · R. D'Ovidio (✉)
Dipartimento di Agrobiologia e Agrochimica,
Università della Tuscia, Via San Camillo de Lellis,
s.n.c, 01100 Viterbo, Italy
e-mail: dovidio@unitus.it

A. Cenci
UMR Diversité et Adaptation des Plantes Cultivées,
Montpellier SupAgro – INRA – IRD – UMII, 2,
Place Pierre Viala, 34060 Montpellier Cedex 1, France

Introduction

Polygalacturonase-inhibiting proteins (PGIPs) are plant cell wall glycoproteins that inhibit fungal and insect *endo*-polygalacturonases (PG) and modulate their activity favoring the accumulation of oligogalacturonides able to induce plant defence responses (De Lorenzo et al. 2001).

PGIP or PGIP-like proteins are also involved in flower development (Gamboa et al. 2001; Jang et al. 2003) and possess anti-freezing properties (Worrall et al. 1998).

PGIPs, like the products of many resistance genes, belong to the subclass of proteins containing leucine-rich repeats (LRRs) of the extracytoplasmic type (Jones and Jones 1997). They typically contain ten imperfect LRRs of 24 residues, each containing the consensus xxLxLxx motif composed of hydrophobic structural residues (L) and by non-conserved residues (x) responsible for PG recognition (Di Matteo et al. 2003).

Plant genomes contain a small number of *Pgip* genes that have been characterized in a number of species (De Lorenzo et al. 2001; D'Ovidio et al. 2004a). These analyses demonstrated that the different members of a *Pgip* family can be quite divergent and can undergo a different transcript regulation during growth and development or following biotic and abiotic stresses. The encoded products can possess different inhibiting activities against fungal or insect PGs and this capability can be affected by single amino acid substitutions and small insertions/deletions (indels) within

the xxLxLxx motifs (Leckie et al. 1999; D'Ovidio et al. 2004b).

Recently, we characterized two expressed pgip genes from *Triticum aestivum* (genome AABBDD), *Tapgip1* and *Tapgip2*, localized on the short arm of homeologous chromosome group 7B and 7D, respectively, but no pgip amplicon or clear hybridization fragments from the A genome were detected (Janni et al. 2006).

In order to investigate the reason for this absence, whether due to gene loss or to a marked sequence rearrangement that strongly reduced the capability of the probe to detect the target sequence, we characterized the pgip genes from the A genome of wild and cultivated wheats. We report also the map location of the *pgip* locus and the identification of *Mutator* transposable elements interrupting the *Tdipgip1* gene of *T. turgidum* ssp. *dicoccoides*.

Materials and methods

Plant materials

Seeds were surface sterilized by immersion in sodium hypochlorite (0.5% v/v) for 30 min, and then rinsed thoroughly in sterile water. Plants were grown at 18–23°C with a 14 h light period. Accessions were kindly provided by National Small Grains Collection, USDA-ARS, Aberdeen, ID, USA (PI), Istituto di Germoplasma, CNR, Bari, Italy (MG), Zentralinstitut fur Genetik und Kulturpflanzenforschung, Gatersleben, Germany (AW and ATRI), International Center for Agricultural Research in the Dry Areas (ICARDA) (ICWT), Istituto Sperimentale per la Cerealicoltura, Lodi, Italy (ID) and researchers (see acknowledgments). The original codes of the accessions are reported in Table S1 in Electronic supplementary material and include the following species: *Triticum urartu* (genome A); *T. monococcum* (A^m) (sspp. *aegilopoides* and *monococcum*); *T. turgidum* (BA) (sspp. *dicoccoides*, *durum*, *dicoccum*, *turanicum*, and *carthlicum*); *T. timopheevii* (GA) (sspp. *armeniacum* and *timopheevii*); *T. aestivum* (BAD) (sspp. *aestivum*, *spelta*, *sphaerococcum*, *macha*, *compactum*); *T. zhukovskyi* (GAA^m).

Chromosome assignment was carried out by using nullitetrasomic and ditelosomic lines of *T. aestivum* cv. Chinese Spring (Sears 1966).

DNA and RNA analyses

Genomic DNA was extracted from 0.15 g of green material following the procedure reported Tai and Tanksley (1991). Plasmid DNA was purified from recombinant bacterial clones with the "NucleoBond® Plasmid Purification Kit" (Clontech, Mountain View, USA).

PCR reactions were performed using the RedTaq™ ReadyMix (Sigma-Aldrich, Milano, Italy) in a 50 µl final volume and following the manufacturer's procedure. Amplification conditions were: 30 cycles at 94°C for 1 min, 65°C for 1 min, 72°C for 1 min and a final step at 72°C for 5 min. Oligonucleotides used as primers were based on *Tapgip1*, *Tapgip2* and on the new reported *Tapgip3* gene and have the following sequences: (PGIPWH1F) 5′-ATGAGCACTCCACCCTCGGCACAT; (PGIPWH1R) 5′-TTATTTCTTGCATGGATCTGGGAG; (PGIPWH4R) 5′-TAGAGGTTGAGGACGAGAAGC; (PGIPWH5F) 5′-TGCCGCAGGTTCGTTACAT; (PGIPWH10R) 5′-GGCGTCGTTGCAAGTGAT; (PGIPWH11R) 5′-GGCGTCGTTGCAAGTGTC; (PGIPWH12R) 5′-GGCGTCGTTGCAAGTGG; (PGIPWH18F) 5′-AACCAAGATGGCATGATACC; (PGIPWH21R) 5′-GTAAAAGTTGAGATCAAGGGG; (PGIPWH5R) 5′-TGCCGCAGGTTCGTTACAT; the specificity of the primers for *Tapgip1* or *Tapgip2* was previously reported (Janni et al. 2006) whereas that for *Tapgip3* was verified in separate PCR experiments using, as a control sample, the BAC clone containing this gene (data not shown). The identity of the PCR product was also confirmed by direct nucleotide sequencing of the amplicon.

Amplified products were fractionated on 1.2% agarose gel, recovered using the QIAquick PCR and QIAquick Gel Extraction Kit (Qiagen srl, Milano, Italy) and directly subjected to nucleotide sequencing.

Sequencing reactions were performed using the "ABI PRISM dye terminator cycle sequencing ready reaction" kit and DNA sequences were determined with the semiautomatic ABI PRISM 310 sequencer (Applied Biosystem, Monza, Italy) and through the sequencing services at MWG-Biotech AG (Ebersberg, Germany).

BLAST (McGinnis and Madden 2004) analyses were performed on the triticeae repetitive (TREP) (http://wheat.pw.usda.gov/ggpages/ITMI/Repeats/index.shtml) and International Nucleotide Sequence Databases (http://www.ncbi.nlm.nih.gov/) databases using the default parameters. A self BLAST analysis was also performed with Blast 2 sequences (Tatusova and Madden 1999) at http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi, using parameters specified in the "Results" section.

Sequence analyses were performed using the DNAMAN software (Lynnon Biosoft, QC, Canada) and sequence alignment were also performed using Clustal W (EMBL-EBI) with the default parameters. Distance matrices were generated by DNAMAN program by using the Neighbor-Joining (Saitou and Nei 1987) method.

A Bacterial Artificial Chromosome library of Langdon#65 (*T. turgidum* ssp. *durum*) (Cenci et al. 2003) was screened by PCR as previously reported (Cenci et al. 2004).

Southern blot analyses were performed as reported in Janni et al. (2006) by using probes labeled with digoxigenin

(Digoxigenin-11-uridine-5′-triphosphate, Roche Diagnostics). Probes were the complete coding region of *Tapgip1* or specific portion of the N- or the C-regions of *Tapgip1*. Probes for the N- or the C-terminal portions were developed by PCR amplifying the first 320 bp (1–320 bp) or the final 146 bp (869–1,014 bp). The specificity of the probes was verified by Southern blot using as target DNA the amplified products corresponding to the N-terminal (320 bp) or C-terminal (146 bp) regions. No cross-hybridization between the two amplicons was observed (data not shown).

Total RNA was extracted using RNeasy Plant Mini kit (Qiagen SpA, Milano, Italy) following the manufacturer's procedure. Contaminating DNA was removed using the DNA-*free*™ mix (Ambion Ltd., Huntingdon, UK) and RNA concentration was determined spectrophotometrically.

RT-PCR experiments were performed by using the QuantyTect® SYBR® Green RT-PCR (Qiagen) in a MyCycler™ thermal cycler (Bio-Rad Life Science, Segrate, Italy), as reported in Janni et al. (2006). Two negative controls missing total RNA template or reverse transcriptase were included in each experiment. Oligonucleotide primers specific for *pgip1* or *pgip3* were those reported above. Amplification of the wheat actin gene (accession number AB181991) was used as reference for transcript amplification and the primers used have the following sequence: (TaACTINF) 5′-AAGAGTCGGTGAAGGGGACT and (TaACTINR) 5′-TTCATACAGCAGGCAAGCAC.

Results

Searching for pgip sequences in the A genome of wild and cultivated wheat

In order to verify the presence of pgip sequences in the A-genome of wheats, we performed Southern blot analysis of genomic DNA from the wild wheat progenitor of the A genome, *T. urartu*, and its close relative, *T. monococcum*. The genomic DNA of the *T. monococcum* digested with *Eco*RI and probed with *Tapgip1* showed a single hybridizing fragment of 4.7 kbp, whereas the accessions PI428315 of *T. urartu* showed two faint hybridizing fragments of 2.9 and 3.7 kbp (Fig. 1, Table 1).

The presence of pgip hybridizing fragments in both *T. urartu* and *T. monococcum*, prompted us to attempt the isolation of the corresponding genes by performing PCR analyses using the primer combination PGIPWH1F/PGIPWH1R, spanning the complete coding region of both *Tapgip1* and *Tapgip2* (Janni et al. 2006). By using these primers we did not obtain amplicons from any of the 16 accessions of *T. urartu* analyzed. On the contrary, we obtained a single amplicon of about 1 kbp from all accessions of *T. monococcum* ssp. *aegilopoides* (ten accessions) and *T. monococcum* ssp. *monococcum* (ten accessions). An example of these results is reported in Fig. 2a, which includes the genotypes analyzed also by Southern blot.

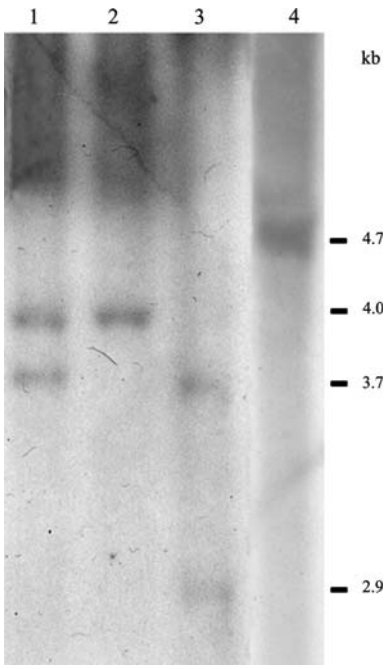

Fig. 1 Hybridization patterns of Pgip sequences in wild and cultivated wheat. Southern blotting of genomic DNA (10 µg) digested with *Eco*RI and probed with *Tapgip1* labeled with digoxigenin. *1 T. aestivum* cv. Chinese spring; *2 T. turgidum* ssp. *durum* cv. Langdon; *3 T. urartu* PI428315; *4 T. monococcum* PI 538722

Table 1 Genetic distances based on the nucleotide sequences of the 5′ and 3′ LTRs interrupting *pgip3*

TU	0				
DIC	0.018	0			
LDN	0.024	0.024	0		
CS	0.023	0.020	0.011	0	
TIM	0.035	0.032	0.035	0.030	0

TU T. urartu, *DIC T. turgidum* ssp. *dicoccoides* MG4343, *LDN T. turgidum* ssp. *durum cv.* Langdon, *CS T. aestivum* cv. Chinese spring, *TIM T. timopheevii*

Direct nucleotide sequencing of the amplicon obtained from the *T. monococcum* ssp. *monococcum* accession ID 1327 proved to be a *Pgip* gene. This gene, named *Tmpgip1* (accession number AM180658), is 1,008 bp and, as the majority of the pgip genes, does not contain introns The product encoded by this gene is composed by 335 amino acids and shows the peculiar PGIP features, including a signal peptide of 26 amino acids, the eight cysteine residues at conserved positions, the ten LRRs, five putative glycosylation sites and about 95% of sequence identity to TaPGIP2 (data not shown).

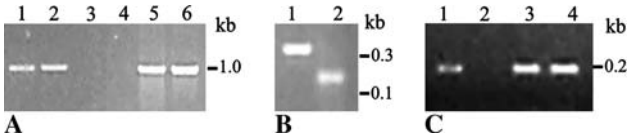

Fig. 2 Agarose gels of pgip amplicons from the A genome. **a** 1.2% agarose gel of PCR amplicons obtained using the primer pair PGIPWH1F/PGIPWH1R spanning the complete coding region of pgip genes. *1 T. aestivum* cv. Chinese spring; *2 T. turgidum* ssp. *durum* cv. Langdon; *3 T. urartu* PI 428315; *4 T. urartu* MG 26992; *5 T. monococcum* PI 538722; *6 T. monococcum* ID 1327. **b** 1.6% agarose gel of pgip amplicons obtained from *T. urartu* PI 428315 with the primer pair PGIPWH1F/PGIPWH4R (*1*), spanning about the first 320 bp of the coding region, and with PGIPWH5F/PGIPWH1R (*2*), spanning 146 bp of the final part of the coding region. **c** Chromosomal assignment of *Tapgip3*. 1.6% agarose gel of PCR products obtained with the primer pair PGIPWH1F/PGIPWH12R. The lack of amplification product indicates the specific chromosomal assignment. *1 T. aestivum* cv. Chinese spring; *2* N7AT7D; *3* N7BT7A; *4* N7DT7A

The genomic DNA of *T. urartu* were subjected to PCR analysis by using additional primer combinations developed from conserved regions along the coding region of both *Tapgip1* and *Tapgip2*. We tested six different primer combinations but only the primer combinations PGIPWH1F/PGIPWH4R, spanning the first 320 bp (from 1 to 320 bp of *Tapgip1*) and PGIPWH5F/PGIPWH1R, spanning the final 146 bp (from 869 to 1,014 bp of *Tapgip1*) produced specific amplicons (Fig. 2b). Sequence analysis of these amplicons demonstrated an high sequence identity (>90%) with *Tapgip1* and *Tapgip2*.

To verify whether the characterized pgip sequences of *T. urartu* were also present in the A genome of cultivated wheats, we developed primers specific for the N-terminal portion of the pgip sequence of *T. urartu* by comparing it with the corresponding ones of *Tapgip1* and *Tapgip2*. PCR analysis on genomic DNA of *T. aestivum* cv. Chinese Spring and its corresponding nulli-tetrasomic lines with the primer pair PGIPWH1F/PGIPWH12R, spanning the first 222 bp (from 1 to 222 bp) of the *T. urartu* pgip sequence, produced a single amplicon, assigned to chromosome 7A (Fig. 2c). Sequence analysis of this amplicon from *T. aestivum* cv. Chinese Spring showed only a few nucleotide substitutions with the corresponding region of *T. urartu*. Taken together these results confirm the existence of a pgip gene in the A-genome of wheat.

Identification of durum wheat BAC clones containing pgip genes

On the basis of the above PCR results, we hypothesized that the pgip gene from the A genome of *T. urartu* and polyploid wheats contains structural features that prevents the amplification of the entire coding region. In order to verify our hypothesis, we screened by PCR a Langdon#65 (*T. turgidum* ssp. *durum*) Bacterial Artificial Chromosome library (Cenci et al. 2003, 2004) by using the conserved primer pair PGIPWH1F/PGIPWH1R, spanning the complete coding region of wheat pgip genes, and the A-genome specific primer pair PGIPWH1F/PGIPWH12R. The screening performed by using the primer pair PGIPWH1F/PGIPWH1R identified the BAC clone BACLDN6526E20 (hereafter BAC26), whereas the primer combination PGIPWH1F/PGIPWH12R identified the BAC clone BACLDN65893O10 (hereafter BAC893).

The two selected BAC clones were subjected to PCR analysis with the conserved primer pairs PGIPWH1F/PGIPWH4R and PGIPWH1F/PGIPWH1R. By using the former primer combination both clones produced the expected fragment of 320 bp, whereas the latter one produced the expected 1 kbp amplicon only from the BAC26. The BAC clone BAC893 did not produced any amplification product (data not shown).

We performed also PCR analysis on both BAC clones using the primer pairs PGIPWH1F/PGIPWH10R (Janni et al. 2006) and PGIPWH1F/PGIPWH12R specific for the pgip genes from the B and A genomes, respectively, and demonstrated that the BAC clones BAC893 and BAC26 harbor pgip genes from the A or B genomes, respectively (data not shown).

The two BAC clones were also subjected to Southern blot analysis, using *Tapgip1* as probe. BAC26 showed the expected 4.0 kb *Eco*RI fragment, whereas BAC893 revealed an hybridization fragment higher than 20 kbp. Southern blot analysis of BAC893 using additional restriction endonucleases that do not cut within the coding region of *Tapgip1* or *Tapgip2*, showed a single hybridization fragment larger than 20 kbp (*Bcl*I, *Apa*I, *Not*I, *Pvu*I, *Sac*I, *Spe*I) or two hybridization fragments (*Eco*RV, *Bam*HI, *Hin*dIII, *Stu*I *Xba*I, *Xho*I). An example of these results is reported in Fig. 3a.

In order to ascertain whether the BAC893 contains two copies of the pgip gene or whether the two hybridization bands derived from different portion of a single copy pgip gene, we performed Southern blot analysis using as probe sequences specific for the N-terminal or the C-terminal portion of the *Tapgip1* coding region. The BAC26 showed the same hybridization fragment with the probe specific for the N- or C-terminal portion, whereas the BAC893 showed hybridization fragments of different size when hybridized with the two probes (Fig. 3). For example, the probe specific for the N-terminal portion produced a single *Bam*HI or *Hin*dIII hybridization fragments of 6.4 or 10.0 kbp, respectively (Fig. 3b, lanes 2), whereas the same Southern blot hybridized with the C-terminal probe produced a single *Bam*HI or *Hin*dIII hybridization fragments of 7.3 and 11.0 kbp, respectively (Fig. 3c, lanes 2).

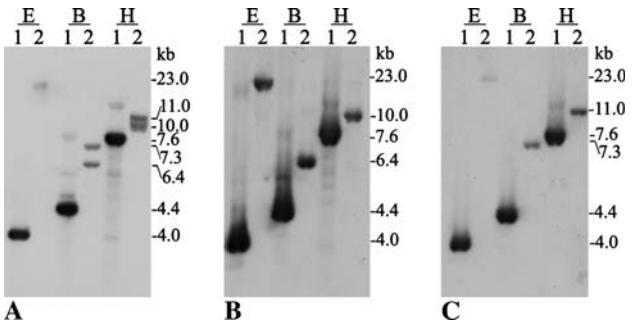

Fig. 3 Hybridization patterns of BAC clones containing *pgip* genes. Southern blots of BAC DNAs (1 μg) digested with *Eco*RI (*E*), *Bam*HI (*B*) and *Hin*dIII (*H*), and hybridized with probes corresponding to the complete coding region (**a**) or specific for the N- (**b**) or C- (**c**) terminal portions of *Tapgip1*. *1* BAC26; *2* BAC893

These results confirmed the hypothesis that only one copy of pgip gene is present in the BAC893 and suggested that the gene is interrupted by a large insert harboring several restriction sites.

Both BAC clones were then subjected to direct nucleotide sequencing. We found that the BAC26 contains a pgip gene, named *Tdpgip1* (accession number AM886139), identical to *Tapgip1*, whereas the BAC893 contains a pgip gene interrupted by an insertion at the beginning of the fourth LRR. The putative coding region of this pgip gene, named *Tdpgip3* (accession number AM884190), shares about 90% nucleotide sequence identity with *Tdpgip1* and contains a single bp deletion at position 292 bp, upstream of the insertion site.

BLAST analysis of the partial sequence of the insertion (750 and 786 bp of the 5′ and 3′ portions, respectively), performed on the TREP database (http://wheat.pw.usda.gov/ggpages/ITMI/Repeats/index.shtml) and on the International Nucleotide Sequence Databases (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi) showed significant similarity only to the barley retrotransposon [long terminal repeat (LTR), *copia*] *Ikeros* [AY853252 (3e-48) and 102J11-1 (1e-19)] and to an EST sequence of *T. monococcum* [BQ803304 (90% identity)]. Detailed sequence comparison with the *Ikeros* elements and analysis of the structural features (primer binding site and poly-purine tract) indicated that the retrotransposon element is oriented in the opposite direction with respect to the coding region of *Tdpgip3*. The size of the LTRs of this retrotransposon element are 373 bp (5′ LTR) and 367 bp (3′ LTR) in the same orientation and flanked by a 5 bp duplicated target site (Fig. 4). The LTRs share 92% sequence identity and the longer size of the 5′ LTR is due to two single bp indels and to four additional base pairs at the end of the sequence that are absent in the 3′ LTR.

RT-PCR on total RNA extracted from seedling of *T. turgidum* cv. Svevo at two-leaf stage (Zadoks stage 12) was carried out to verify if *Tdpgip3* is transcriptionally active. Analyses performed by using the primer combination PGIPWH1F/PGIPWH12R specific for *Tdpgip3* did not produced any amplification product. On the contrary, the primer combination PGIPWH1F/PGIPWH10R specific for *Tdpgip1* gave the expected amplicon of 219 bp (Fig. 5).

Incidence and variability of the retroelement interrupting *Tdpgip3*

In order to verify the occurrence of the *copia* retroelement interrupting *Tdpgip3* in the A genome of durum wheat Landon#65, we analyzed by PCR a number of wild and polyploidy wheat genotypes. To perform this analysis we developed two primer pairs whose amplification products span the insertion site at both the 5′ or 3′ ends of the *copia* retroelement. The primer pair PGIPWH1F/PGIPWH21R amplified 850 bp spanning the N-terminal portion of the *Tdpgip3* and the complete 3′ LTR of the *copia* retrotransposon, whereas the primer pair PGIPWH18F/PGIPWH5R amplified 823 bp spanning the complete 5′ LTR and the C-terminal portion of *Tdpgip3*. We performed this analysis in a number of wheat species, specified in the "Materials and methods", and all genotypes but *T. monococcum* accessions produced the expected amplicons of 850 and 823 bp (not shown).

In order to assess the sequence conservation of this retroelement, we sequenced both the above specified LTR amplicons from *T. urartu* PI 428315 (*Tupgip3*, accession number AM884191), *T. turgidum* ssp. *dicoccoides* MG4343 (*Tdipgip3*, accession number AM884193), *T. timopheevii* PI 221421 (*Ttpgip3*, accession number AM884192), *T. aestivum* cv. Chinese Spring (*Tapgip3*,

```
     |     LRR pgip3  |       3′ LTR     copia retroelement   5′  LTR     |    LRR  pgip3  |
LDN  CTGACCAAGCTCGAAGTGATAGATCCTATGGATCGTGGT--//--GGTGGAGAGATCAACCTAATACAGAGAACTCACCCTCTCCTC
TU   CTGACCAAGCTCGGAGTGATAGATCCTATGGATCGTGGT--//--GGTGGAGAGATCAACCTAATACAGAGAACTCACCCTCTCCTC
DIC  CTGACCAAGCTCGGAGTGATAGATCCTATGGATCGTGGT--//--GGTGGTGAGATCAACCTAATACAGAGAACTCACCCTCTCCTC
TIM  CTGACCAAGCTCGGAGTGATAGATCCTATGGATCGTGGT--//--GGTGGAGAGATCAACCTAATACAGAGAACTCACCCTCTCCTC
CS   CTGACCAAGCTCGGAGTGATAGATCCTATGGATCGTGGT--//--GGTGGTGAGATCAACCTAATACAGAGAACTCACCCTCTCCTC
     ************ ************************      ***** ************************************
```

Fig. 4 Nucleotide comparison between the different *pgip3* genes at the insertion site of the retroelements. The *copia* retrotransposons interrupt the coding regions at the same position (at the beginning of the fourth LRR) and show conserved sequence at the insertion site. *LDN T. turgidum* ssp. *durum* cv. Langdon, *TU T. urartu* PI 428315, *DIC T. turgidum* ssp. *dicoccoides* MG4343, *TIM T. timopheevii* PI 221421, *CS T. aestivum* cv. Chinese Spring. In *gray* is the duplicated target site

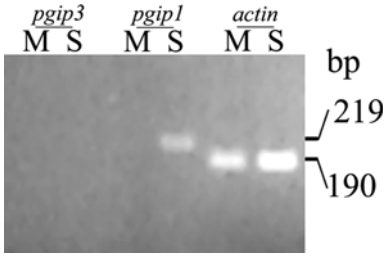

Fig. 5 Transcript analysis of pgip genes in wheat plants. RT-PCR was performed on total RNA using gene-specific primers and the amplification products fractionated on 2% agarose gel. *pgip3* RT-PCR with primer pair PGIPWH1F/PGIPWH12R, specific for *pgip3*; *pgip1* RT-PCR with primer pair PGIPWH1F/PGIPWH10R, specific for *pgip1*; *actin* RT-PCR with primer pair TaACTINF/TaACTINR, specific for a wheat *actin* gene (control). *M T. turgidum* ssp. *dicoccoides* MG4343, *S T. turgidum* ssp. *durum* cv. Svevo. Both *pgip1* and *pgip3* transcripts are missing in *T. turgidum* ssp. *dicoccoides* MG4343, whereas only *pgip3* transcript is missing in *T. turgidum* ssp. *durum* cv. Svevo

accession number AM884194). Nucleotide sequence comparison between the combined 5′ and 3′ LTRs of each of these genotypes, including *T. turgidum* cv. Langdon, showed reduced sequence polymorphism (>96.5% identity) and that *T. timopheevii* has the most distant sequence (Table 1). We have used the available LTR sequences to estimate also the time of insertion using the synonymous substitution rate of $1.3E^{-8}$ which was proposed by Ma and Bennetzen (2004). All estimations were consistent with an insertion time of 1.9 million years ago (MYA).

Genetic mapping of the *Pgip* locus

In order to map the *Pgip* locus on the *T. turgidum* genome, we explored the polymorphism between the parental tetraploid genotypes *T. turgidum* ssp. *durum* cv. Messapia and *T. turgidum* ssp. *dicoccoides* MG4343, the parental genotypes of a recombinant inbred lines (RILs) mapping population (Blanco et al. 2004). PCR performed with the primer combination PGIPWH1F/PGIPWH1R on Messapia and MG4343 produced amplicons of different size, 1 and 1.5 kbp, respectively. The 65 RILs derived from Messapia and MG4343 were amplified with the same primers combination and scored for the size of the amplified fragment. The segregation data integrated on the durum wheat map (Blanco et al. 2004) indicated that *Pgip* maps on the short arm of chromosome 7B, 17.5 cM from the centromere between XS13M50(330) and Xmgb105s [XS13M50(330)-5.0-PGIP-11.7-Xmgb105s].

Characterization of the *Tdipgip1* gene of *T. turgidum* ssp. *dicoccoides* MG4343

In order to characterize the origin of the larger size of the pgip amplicon obtained from the accession MG4343, we performed Southern blot analysis and a direct nucleotide sequencing of the 1.5 kb amplicon. Southern blot analysis, using *Tapgip1* as probe, showed a single hybridization fragment of about 4.5 kbp, in place of the 4.0 kbp of durum and bread wheat (Fig. S3), whereas the nucleotide sequence of the amplicon showed an almost perfect identity (only two synonymous and one non-synonymous nucleotide substitutions) with *Tapgip1*, except for the presence of a 463 bp insertion within the sixth LRR that interrupts the coding region of this gene. This interrupted gene was named *Tdipgip1a* (accession number AM884195).

The nucleotide sequence of the insertion was submitted to BLAST analysis with the TREP database but no significant similarity were found. On the contrary, when the BLAST analysis was performed at the International Nucleotide Sequence Databases (nr, est, gss and htgs) several significant similarity (E < 1e-3) were found (Table S2).

A self BLAST analysis [Blast 2 sequences, Tatusova and Madden (1999)] showed also the insert sequence has palindromic structure in the central region that can be extended to the ends when less stringent parameter (match/mismatch reward equal to 2/−3 and 3/−4) were used.

A perfect 10 bp repeat (GACCGAGTTC) flanks the insert, one repeat belonging to *Tdipgip1a*, the other belonging to the insert.

Due to (1) the repetitive nature of the insert, (2) the palindromic structure of its sequence and (3) the 10 bp repeat at its ends (duplication of insertion site), it was concluded that the insertion is a Class II transposable element, belonging to the *Mutator* superfamily.

Since no similar sequences were found in the TREP database, nor annotations were available for the other copies found in the sequence databases, the insert was considered as a new element and was called *Vacuna*.

Incidence and variability of the *Vacuna* element interrupting *Tdipgip1*

To verify the occurrence of the *Vacuna* element, we analyzed the *pgip1* in the same diploid and polyploidy wheat genotypes analyzed for the *copia* retrotransposon. The analysis was performed by using the primer pair PGIPWH1F/PGIPWH1R and in all cases we obtained the amplicon of the expected size (1 kbp), except in the accession PI466960 of *T. turgidum* ssp. *dicoccoides* that showed an amplicon of 1.5 kbp. Nucleotide sequence of this amplicon showed the presence of an insertion at the beginning of the gene (named *Tdipgip1b*, accession number AM884196), within the region encoding the signal peptide. The nucleotide sequence of the insert show a perfect 9 bp repeat (TCGTCCTCC) flanking the insert, one repeat belonging to *Tdipgip1b*, the other one to the insert. Nucleotide sequence

comparison of the insertion showed about 80% identity to the *Vacuna* transposon and values ranging between 72 and 80% with the other sequences listed on Table S2 (Fig. 6). Blast searches did not identify any additional sequence with similarity values close or higher than those obtained with the sequence listed on Table S2. On this basis, the alignment of all these elements was produced and the majority consensus sequence derived (Fig. 6). Since the sequence identity between the consensus and any element was always higher than 80%, all these elements were classed in the *Vacuna* family, as recommended by Wicker et al. (2007).

Finally, we verified that *Tdipgip1a* is transcriptionally inactive by performing RT-PCR on total RNA extracted from seedling of the MG4343 accession. Moreover, since this genotype possess the AABB genome, we further verified that also its *pgip3* was inactive. Because of the conservation of the target regions, we performed the RT-PCR by using the same gene-specific primers used above. As expected, the analyses performed by using the primer combinations PGIPWH1F/PGIPWH10R, specific for *pgip1*, or PGIPWH1F/PGIPWH12R, specific for *pgip3*, did not produced any amplification product (Fig. 5).

Discussion

In this study, we have shown that the A-genome of cultivated wheat contains a pgip gene (*pgip3*) that is interrupted by an LTR, *copia* retrotransposon insertion. Our results showed also that this insertion is present in the *pgip3* of *T. urartu* (AA) and in all the polyploidy wheats assayed, including *T. turgidum* (BBAA) and *T. timopheevii* (GGAA), but is absent in the *pgip3* of *T. monococcum* (A^mA^m). On the basis of these results it was concluded that the retrotransposon insertion took place after the divergence between *T. monococcum* and *T. urartu* but before the formation of the polyploid wheats, with an estimations of the insertion time of 1.9 MYA. This finding is slightly higher than the estimation of the divergence time for the wheat A genomes of 0.5–1.5 MYA, when their date is re-estimated with the 1.3×10^{-8} substitution rate (Wicker et al. 2003) and represents an additional evidence that the A genomes of *T. turgidum*, *T. timopheevii* and *T. aestivum* derived from the A genome of *T. urartu*, as previously reported (Dvorak et al. 1993).

Moreover, sequence comparison between the retrotransposons confirms previous indication that *T. urartu* genome appears more similar to the A genomes of *T. turgidum* and

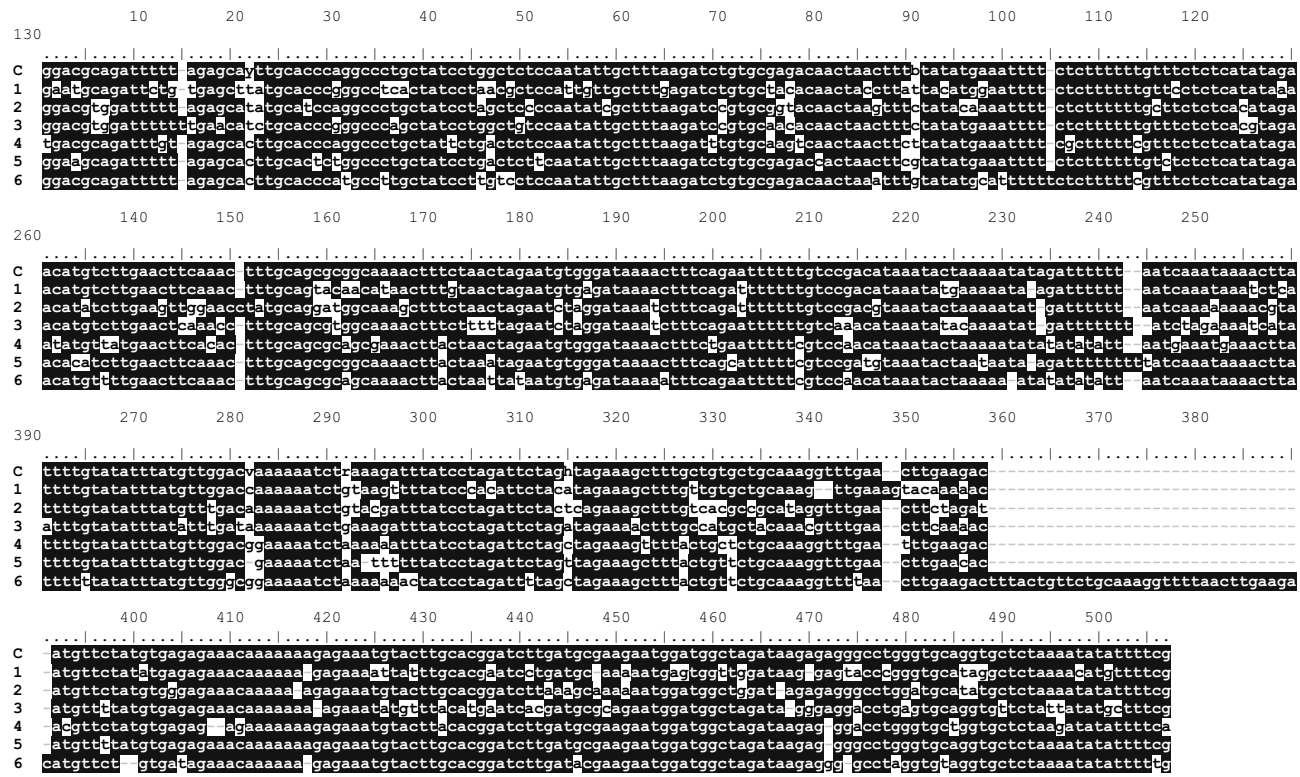

Fig. 6 Sequence alignment between *Vacuna* elements (without duplication of insertion site) interrupting *Tdipgip1a* (*1*), *Tdipgip1b* (*2*) and those complete found in the databases: CT009585 and CT009588 (*3*), AF509533 (*4*), AF509534 (*5*), and CW991740 (*6*); (*C*) represents the derived majority consensus sequence

T. aestivum than the A genome of *T. timopheevii* (Khlestkina and Salina 2001; Cenci and David, unpublished data).

Most plant species, especially those with a large genome, such as wheat, contain a large quantities of transposable elements. These sequences reside in the vast expanses between genes but also within or near genes and consequently can have a role in the evolution of gene structure and expression (for reviews see Wicker et al. 2007; Feschotte et al. 2002). The LTRs of the retroelement found within the fourth LRR of the *pgip3* gene show a high sequence identity (>70%) to the barley retrotransposon (LTR, copia) *Ikeros* [AY853252 (3e-48) and 102J11-1 (1e-19)]. The *copia* retrotransposons are major components of all plant genomes and members of the *Ikeros* family have been identified, so far, only in barley and rice (*Ostonor1*) (Wicker and Keller 2007). Our finding extends the *Ikeros* family to wheat and consequently represents an additional evidence of the conservation of *copia* elements within the *Triticeae* genomes.

Based on Southern blots and sequence analyses of BAC26 and BAC893, wheat contains one copy of the pgip gene per genome. In addition to the inactive *Tapgip3* (genome A), *T. aestivum* contains the constitutively expressed *Tapgip1* (genome B) and *Tapgip2* (genome D) (Janni et al. 2006). These genes, including also the *pgip3* of *T. monococcum* (A^mA^m), have a very high level of sequence conservation (>92%), and their encoded proteins show the typical structure composed of ten LRRs. Until now, no pgip genes interrupted by transposons have been reported, although insertion of *copia* retroelements into the coding region of two LRR members of the rice *Xa21* resistance gene family have been identified (Song et al. 1997; Wang et al. 1998). In these two cases, since the insertions generated putative truncated proteins with new functional features, it has been suggested that retrotransposons can be important players for diversification of this gene family (Song et al. 1997; Wang et al. 1998). The absence of *pgip3* transcripts in wheat tissues indicates that no truncated PGIP3 protein is produced. Nevertheless, the retrotransposon insertion into *pgip3* of *T. urartu* and its conservation in the A genomes of polyploidy wheat species indicates that the lack of this component does not have a great impact on the fitness of these genotypes. This possibility might be related to the occurrence of PGIP redundancy in wheat tissue. In fact, in addition to the constitutively expressed *Tapgip1* and *Tapgip2*, bread wheat contains PG-inibiting proteins with N-terminal sequences different from the classical ones (Lin and Li 2002; Kemp et al. 2003). A similar complex pattern of PG-inhibiting proteins should be also expressed in *T. urartu* because, in spite of the presence of only the inactive *pgip3*, its leaf tissue shows PGIP activity (M. Di Giovanni and R. D'Ovidio, unpublished data).

In addition to the retrotransposon insertion that inactivated the *pgip3* gene, we found a Class II transposable element, that we named *Vacuna*, interrupting *pgip1* in two accessions of *T. turgidum* ssp. *dicoccoides*. The two transposons represent new elements belonging to the *Mutator* superfamily and since they interrupt the coding region of *Tdipgip1* in different positions, it was concluded that they originated by independent insertion events. This finding and the occurrence of two additional insertion events of *Vacuna* elements within the 3′ flanking region of linked LRR genes of *T. tauschii* (AF509533 and AF509534) suggests that *Vacuna* might move preferentially within short distances, as showed for the *Ac/Ds* transposons (Bancroft and Dean 1993; Keller et al. 1993).

The insertion of the *Vacuna* elements might be responsible for the transcriptional inactivation of the *Tdipgip1* genes. As a result, the MG4343 and PI 466960 accessions of *T. turgidum* ssp. *dicoccoides* do not have any functional pgip gene possessing the typical structure, being also *pgip3* silenced by the *copia* retrotransposon. However, both accessions of *T. turgidum* ssp. *dicoccoides* possess PG-inhibiting activity (C. Volpi and R. D'Ovidio, unpublished data), possibly due, as already mentioned above, to PG-inhibiting proteins with sequences different from the classical one.

In conclusion, we demonstrated that the durum wheat genome contains a single copy pgip gene per genome, localized in the pericentric region of the short arm of chromosome group 7. We showed, for the first time to our knowledge, that pgip genes experienced inactivation by transposable elements. In particular, the inactivation of *pgip3* by a *copia* retrotransposon insertion is conserved in the A genome of wheats since the divergence between *T. urartu* and *T. monococcum*, whereas the inactivation of *Tdipgip1* by the new transposable elements *Vacuna* occurred at least twice. We speculate that this inactivation are tolerated because of redundancy of PGIP activities in the wheat genome.

Acknowledgments Research supported by MIUR [Ministero dell'Istruzione, dell'Università e della Ricerca; grants PRIN (Programmi di Ricerca Scientifica di Rilevante Interesse Nazionale) 2005 and FISR (Fondo Integrativo Speciale per la Ricerca) 2005–2008]. We thank Profs. Domenico Lafiandra and Mario Augusto Pagnotta (Università della Tuscia, Viterbo, Italy) for providing seeds of some wheat accessions and cultivars and Dr. Thomas Wicker for assistance in estimating the insertion date of transposon.

References

Bancroft I, Dean C (1993) Transposition pattern of the maize element Ds in *Arabidopsis thaliana*. Genetics 134:1221–1229

Blanco A, Simeone R, Cenci A, Gadaleta A, Tanzarella OA, Porceddu E, Salvi S, Tuberosa R, Figliuolo G, Spagnoletti P, Röder MS,

Korzun V (2004) Extension of the Messapia × dicoccoides linkage map of *Triticum turgidum* (L). Thell Cell Mol Biol Lett 9:529–541

Cenci A, Chantret N, Xy K, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) A half million clones bacterial artificial chromosome (BAC) library of durum wheat. Theor Appl Genet 107:931–939

Cenci A, Somma S, Chantret N, Dubcovsky J, Blanco A (2004) PCR identification of durum wheat BAC clones containing genes coding for carotenoid biosynthesis enzymes and their chromosome localization. Genome 47:911–917

D'Ovidio R, Raiola A, Capodicasa C, Devoto A, Pontiggia D, Roberti S, Galletti R, Conti E, O'Sullivan D, De Lorenzo G (2004b) Characterization of the complex locus of *Phaseolus vulgaris* encoding polygalacturonase-inhibiting proteins (PGIPs) reveals sub-functionalization for defense against fungi and insects. Plant Physiol 135:2424–2435

D'Ovidio R, Mattei B, Roberti S, Bellincampi D (2004a) Polygalacturonases, polygalacturonase inhibiting proteins and pectic oligomers in plant-pathogen interactions. Biochim Biophys Acta 1696:237–244

De Lorenzo G, D'Ovidio R, Cervone F (2001) The role of polygacturonase-inhibiting proteins (PGIPs) in defense against pathogenic fungi. Annu Rev Phytopathol 39:313–335

Di Matteo A, Federici L, Mattei B, Salvi G, Johnson KA, Savino C, De Lorenzo G, Tsernoglou D, Cervone F (2003) The crystal structure of PGIP (polygalacturonase-inhibiting protein), a leucine-rich repeat protein involved in plant defense. Proc Natl Acad Sci USA 100:10124–10128

Dvorak J, Di Terlizzi P, Zhang HB, Resta P (1993) The evolution of polyploid wheats: identification of the A genome donor species. Genome 36:21–31

Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet 3:329–341

Gamboa A, Paez-Valencia J, Acevedo GF, Vazquez-Moreno L, Alvarez-Buylla RE (2001) Floral transcription factor AGAMOUS interacts in vitro with a leucine-rich repeat and an acid phosphatase protein complex. Biochem Biophys Res Commun 288:1018–1026

Jang S, Lee B, Kim C, Yim J, Han J-J, Lee S, Kim S-R, An G (2003) The OsFOR1 gene encodes a polygalacturonase-inhibiting protein (PGIP) that regulates floral organ number in rice. Plant Mol Biol 53:357–369

Janni M, Di Giovanni M, Roberti S, Capodicasa C, D'Ovidio R (2006) Characterization of expressed Pgip genes in rice and wheat reveals similar extent of sequence variation to dicot PGIPs and identifies an active PGIP lacking an entire LRR repeat. Theor Appl Genet 113:1233–1245

Jones DA, Jones JDG (1997) The roles of leucine rich repeats in plant defences. Adv Bot Res 24:90–167

Keller J, Lim E, Dooner HK (1993) Preferential transposition ofAc to linked sites in Arabidopsis. Theor Appl Genet 86:585–588

Kemp G, Bergmann CW, Clay R, Van der Westhuizen AJ, Pretorius ZA (2003) Isolation of a polygalacturonase-inhibiting protein (PGIP) from wheat. Mol Plant Microbe Interact 16:955–961

Khlestkina EK, Salina E A (2001) Genome-specific markers of tetraploid wheats and their putative diploid progenitor species. Plant Breed 120:227–232

Leckie F, Mattei B, Capodicasa C, Hemmings A, Nuss L, Aracri B, De Lorenzo G, Cervone F (1999) The specificity of polygalacturonase-inhibiting protein (PGIP): a single amino acid substitution in the solvent-exposed β-strand/β-turn region of the leucine-rich repeats (LRRs) confers a new recognition capability. EMBO J 18:2352–2363

Lin W, Li Z (2002) The partial structure of wheat polygalacturonase-inhibiting protein. Zhongguo Shengwu Huaxue Yu Fenzi Shengwu Xuebao 18:197–201

Ma J, Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci 101:12404–12410

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 32:W20–W25

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Sears E R (1966) Nullisomic-tetrasomic combination in hexaploid wheat. In: Riley R, Lewis KR (eds) Chromosome manipulation and plant genetics. Oliver and Boyd, Edinburg, pp 29–45

Song W-Y, Pi L-Y, Wang G-L, Gardner J, Holsten T, Ronald PC (1997) Evolution of the rice *Xa21* disease resistance gene family. Plant Cell 9:1279–1287

Tai T, Tanksley S (1991) A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. Plant Mol Biol Rep 8:297–303

Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247–250

Wang G-L, Ruan D-L, Song W-Y, Sideris S, Chen L, Pi L-Y, Zhang S, Zhang Z, Fauquet C, Gaut BS, Whalen MC, Ronald PC (1998) *Xa21D* encodes a receptor-like molecule with a Leucine-rich repeat domain that determines race-specific recognition and is subject to adaptive evolution. Plant Cell 10:765–780

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu Z, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. Plant Cell 15:1186–1197

Wicker T, Keller B (2007) Genome-wide comparative analysis of *copia* retrotransposons in *Triticeae*, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. Genome Res 17:1072–1081

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman A (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982

Worrall D, Elias L, Ashford D, Smallwood M, Sidebottom C, Lillford P, Telford J, Holt C, Bowles D (1998) A carrot leucine-rich-repeat protein that inhibits ice recrystallization. Science 282:115–117