Integrase diversity and transcription of the maize retrotransposon Grande

Eva Gómez, Alan H. Schulman, José Antonio Martínez-Izquierdo, and Carlos M. Vicient

Abstract: Grande is an abundant gypsy-like retrotransposon present in the genera Zea and Tripsacum. Related retrotransposon families can be found in sorghum, rice, and barley, but not in wheat or rye. We have amplified and sequenced several copies of part of the integrase domain derived from the Zea mays, Zea diploperennis, and Tripsacum dactyloides genomes. There are no significant differences in divergence or clustering between the integrase sequences of these species. The substitution rate for synonymous sites was found to be higher than those of non-synonymous sites; this indicates that Grande integrase has been under purifying selection for function. Grande is transcribed in leaves. The transcripts show sequence diversity similar to that of genomic sequences, but belong to restricted clades; this indicates that only some evolutionary branches of Grande have retained transcriptional competence.

Key words: maize, plant genome, retrotransposon, Zea diploperennis, Zea genus.

Résumé : L’élément Grande est un rétrotransposon de type gypsy qui est abondant chez les genres Zea et Tripsacum. Des familles apparentées de rétrotransposons sont présentes chez le sorgho, le riz et l’orge, mais pas chez le blé ou le seigle. Les auteurs ont amplifié et séquencé plusieurs copies d’une partie du domaine intégrase chez les génomes du Zea mays, du Zea diploperennis et du Tripsacum dactyloides. Il n’y a pas de différences significatives quant à la divergence ou au regroupement entre les séquences d’intégrase chez ces espèces. Le taux moyen de substitutions synonymes plus élevé que celui de substitutions non-synonymes indique que l’intégrase de Grande a subi une sélection purificatrice pour sa fonction. L’élément Grande est transcrit dans les feuilles. Les transcrits présentent une diversité nucléotidique semblable à celle observée chez les séquences génomiques, mais appartiennent à des clades restreints, ce qui indique que seules certaines branches évolutives de l’élément Grande ont conservé la capacité d’être transcrites.

Mots clés : maïs, génome de plante, rétrotransposon, Zea diploperennis, genre Zea.

Introduction

Long terminal repeat (LTR) retrotransposons are class I transposable elements whose life cycles include the transcription of integrated copies to RNA, the copying of RNA to cDNA by reverse transcriptase (RT), and the insertion of cDNA into the genome by integrase (IN) (Suoniemi et al. 1998a; Vicient and Schulman 2002). They are ubiquitous and generally abundant in the vascular plants. Transcription is then the first step in retrotransposon replication. Low levels of transcription can be detected for many plant retrotransposons (Vicient et al. 2001), but abundant transcripts have only been found from a few families (Pouteau et al. 1991; Hirochika et al. 1996). Generally, retrotransposons fall into 2 groups: prevalent, transcriptionally active elements (e.g., BARE-1) that integrate into repetitive DNA, and rare, stress- or tissue-culture-activated families (e.g., Tos17) that integrate into genic regions and generate mutations.

Retrotransposon replication is error prone owing to the low fidelity and lack of proofreading activity of RT and
RNA polymerase II. Integrated, replicated copies will therefore display structural diversity and can be viewed as populations of closely related but different copies (Casacuberta et al. 1995). IN function and evolution have been intensively studied for the retroviruses (Hickman et al. 1994), but less is known about the evolution of the IN of retrotransposons (Suoniemi et al. 1998b; Shcherban et al. 2001).

The genus Zea belongs to the family Poaceae and is closely related to the genus Tripsacum (Buckler and Holtsford 1996). In addition to maize (Zea mays), the genus Zea has 3 more species: Zea diploperennis, Zea perennis, and Zea luxurians. Many retrotransposon families have been identified in Zea genomes (Meyers et al. 2001). Among the Zea retrotransposons, the gypsy-like Grande was isolated initially from Z. diploperennis (Martínez-Izquierdo et al. 1997) but is also present in the other Zea species (García-Martínez and Martínez-Izquierdo 2003). Grande, with more than 5000 copies per haploid genome, is one of the most abundant Zea retrotransposons (García-Martínez and Martínez-Izquierdo 2003).

Materials and methods

Plant DNA source and preparation

Plants were grown in a controlled-environment chamber at 28 °C in the dark until they were 5–10 d old. Genomic DNA was extracted from leaves using the DNeasy Plant mini kit (Qiagen, Valencia, Calif.). The following species were studied: Z. mays subsp. mexicana, Z. mays subsp. parviglumis, and Z. mays subsp. mays (‘B73’, ‘W64A’, and ‘Palomero Tolujeño’), Z. perennis, Z. diploperennis, Tripsacum dactyloides, Sorghum bicolor, Oryza sativa (rice), Hordeum vulgare (barley), Triticum aestivum (wheat), and Secale cereale (rye).

PCR primers and reactions

The IN domain was amplified by PCR using the following primers: forward, 5′-TGGGATTTGCTAAGCCATTC-3′; reverse, 5′-GACTCTGGCATTATCGGTTC-3′. Primers were designed based on previously sequenced INs. The reaction mixtures comprised 1 µg genomic DNA as template, 0.2 mmol/L each dNTP, 100 pmol each primer, 2.5 U Taq polymerase in 1× Buffer A (Promega, Madison, Wis., catalogue No. M1861), and water to a total volume of 50 µL. The mixture was overlaid with paraffin oil. PCRs consisted of an initial denaturation step at 94 °C for 2 min; 40 cycles of denaturation at 94° for 30 s, annealing at 56 °C for 45 s, and extension at 72 °C for 2 min; and a final extension step at 72 °C for 2 min. Reactions were carried out in an MJ Research PTC-200 Peltier Thermal Cycler (Waltham, Mass.).

cDNA amplification

Total RNA was isolated from maize leaves using the RNeasy Plant Mini Kit (Qiagen, catalogue No. 74904). RNA was treated with 3 U RNAse-free DNase I (Roche, Basel, Switzerland, catalogue No. 0776785) for 1 h at 37 °C. The cDNA was produced using the One Step RT-PCR Kit (Qiagen, catalogue No. 210210) using the same primers as in the PCR amplifications.

Cloning and sequencing

The reaction products were purified using the Qiaquick PCR Purification Kit (Qiagen, catalogue No. 2870) and cloned into the pGEM-T vector (Promega, catalogue No. A1360). Sequencing reactions on plasmid miniprep were performed with Sequenase v. 2.0 (Amersham Pharmacia Biotech, Uppsala, Sweden, catalogue No. US70750) and analyzed under standard conditions with an automated system. The amplified sequences are deposited in GenBank under accession Nos. AJ634383–AJ634417.

Sequence alignment and phylogenetic analysis

Alignment of the sequences was done with the CLUSTALX program (Thompson et al. 1997) and a neighbor-joining tree was calculated using the TREECON program (Van de Peer and De Wachter 1994). The accuracy of the neighbour-joining tree was evaluated by bootstrap analysis and 1000 bootstrap replicates were calculated. Sequence diversity, K_s and K_v values were calculated using DNAsP v. 4.0 (Rozas et al. 2003).

Results and discussion

The presence of Grande within various members of the Poaceae was assayed in a PCR survey using primers that amplify the N-terminal part of the Grande integrase domain. An abundant fragment of the expected size was detected using genomic DNA from Zea species and Tripsacum (Fig. 1a). Products of the same size, but in much lower yield, were also detected in Sorghum, Oryza, and Hordeum. No products were evident in Triticum or Secale under our PCR conditions. The genus Sorghum is closely related to Zea, but the genome size is much smaller. This difference in genome size is correlated with a general paucity of retrotransposons in Sorghum (Tikhonov et al. 1999), and is consistent with the poor PCR amplification of Grande that we observed in Sorghum. Tripsacum is the most closely related genus to Zea. Many retrotransposons first characterized in maize are also present in Tripsacum (Purugganan and Wessler 1994; Vicent and Martínez-Izquierdo 1997; Marillonnet and Wessler 1998; Meyers et al. 2001). Grande is among these (Monfort et al. 1995; Fig. 1a). We have also been able to amplify sequences related to Grande from rice and barley. In rice, RIRE2, a gypsy-like retrotransposon, is similar to Grande in the gag-pol region (Jiang et al. 2002). Although no complete element resembling Grande has yet been characterized in barley, a sequence related to Grande was amplified by inverse PCR as the insertion site of a BARE-1 element (Suoniemi et al. 1997). The data here, showing that primers designed to Grande IN can amplify products of conserved length from barley (Fig. 1a), implies that the retrotransposon family is old and fairly well conserved. The amplification data indicate that Grande is a member of an ancient superfamily of elements found throughout the Poaceae.

The same primers used to amplify IN from genomic DNA were tested with cDNA by RT-PCR, employing RNA extracted from maize leaves as the template (Fig. 1b). The RT-PCRs amplified a fragment of the expected size, whereas control reactions lacking mRNA in the reverse-transcription step failed to amplify products, thereby indicating that at
least some copies of *Grande* are transcriptionally active in this tissue. Some of the most abundant elements in the genome appear to be transcribed in most tissues, but at low levels (Meyers et al. 2001).

Amplified fragments from maize ‘B73’, *Z. diploperennis*, and *Tripsacum dactyloides* genomes and from maize leaf mRNA were excised from gels, purified, and cloned; independent clones of each sample were then sequenced and aligned (Supplementary data2). Comparing only genomic sequences, the proportion of polymorphic sites was 0.39, which is similar to what was found for *magellan* retrotransposon LTRs (Purugganan and Wessler 1994). The average nucleotide diversity per site as measured by the Jukes–Cantor method (Lynch and Crease 1990) was calculated both between and within samples (Table 1). The levels of within-species diversity were similar for the 2 Zea species and lower for *Tripsacum*. Diversity within maize genomic sequences and within maize cDNAs was similar. Between-species diversity was similar for genomic sequences, but higher for cDNA.

The *Grande* IN domain possesses a level of nucleotide polymorphism 1.26 times higher than the LTR retrotransposon sequence of *magellan* (Purugganan and Wessler 1994), even though LTRs are thought to evolve considerably more rapidly than IN domains (Lankenau et al. 1990; Lyubomirskaya et al. 1990). The difference between the rate of sequence change observed for *Grande* and the error rate for reverse transcription, which is several orders of magnitude greater, likely parallels the relative proportion of replication as a chromosomal component via cellular DNA polymerase rather than through the retrotransposon life cycle. The higher diversity of the maize sequences and of the cDNAs as a whole therefore suggests a higher replicative activity for these of *Grande* elements. The rates of synonymous ($K_s$) and non-synonymous ($K_a$) substitutions were calculated for each group of sequences. $K_s$ was 5.3–7.7 times higher than $K_a$, suggesting that the *Grande* IN sequence has been under functional selective pressure. Interestingly, the *Tripsacum* ($K_a/K_s = 5.7$) and cDNA sequences (5.3) showed lower values within each group.

The sequences were conceptually translated and the predicted proteins were aligned. Seven sequences contained stop codons. Curiously, 5 of them came from mRNA and only 2 came from genomic DNA. The neighbor-joining method was used to predict phylogenetic relationships (Fig. 2). Two general conclusions could be garnered. First, cDNA sequences formed separate groups that are more homogeneous than sequences from genomic DNA. The phylogenetic analysis and diversity measurements together support the independent, LTR-driven transcriptional origin of the RNA sequences, rather than the alternatives of either read-through transcription from cellular promoters or con-

---

2 Supplementary data for this article are available on the journal Web site (http://genome.nrc.ca) or may be purchased from the Depository of Unpublished Data, Document Delivery, CISTI, National Research Council Canada, Building M-55, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada. DUD 5048. For more information on obtaining material refer to http://cisti-icist.nrc-cnrc.gc.ca/irm/unpub_c.shtml.
tamination of genomic DNA in the RNA preparations. Both would tend to randomize the distribution of putative Grande transcripts with respect to the genomic DNA sequences. All this suggests that only some subfamilies of Grande are transcriptionally active. Secondly, genomic sequences from the various species appear to be randomly mixed. The separation of Zea and Tripsacum took place 4.5–4.8 million years ago (Hilton and Gaut 1998) and many of the Grande subfamilies appear to have arisen before this event (SanMiguel et al. 1998). This is supported by the presence of elements related to Grande in phylogenetically distant grasses, such as barley.

Acknowledgements

We thank Anne-Mari Narvanto for her excellent technical assistance. E.G.O. was the recipient of a grant from European Union Biotechnology Program (ERBIO40CT960508). C.M.V. was supported by Academy of Finland Project 44404. This work was supported in part by the following grants: project 53453 (Academy of Finland), PB97-1255 (Direccion General de Investigacion Cientifica y Tecnica, and BIO99-1175 (Comision Interministerial de Ciencia y Tecnologia) from the Spanish Ministerio de Educacion y Ciencia.

References


© 2006 NRC Canada


