# Characterization of High-Copy-Number Retrotransposons From the Large Genomes of the Louisiana Iris Species and Their Use as Molecular Markers

**Edward K. Kentner,[1] Michael L. Arnold and Susan R. Wessler**

*Department of Genetics, University of Georgia, Athens, Georgia 30602*

Manuscript received December 3, 2002
Accepted for publication February 26, 2003

ABSTRACT

The Louisiana iris species *Iris brevicaulis* and *I. fulva* are morphologically and karyotypically distinct yet frequently hybridize in nature. A group of high-copy-number TY3/*gypsy*-like retrotransposons was characterized from these species and used to develop molecular markers that take advantage of the abundance and distribution of these elements in the large iris genome. The copy number of these IRRE elements (for *ir*is *r*etro*e*lement), is $\sim 1 \times 10^5$, accounting for ~6–10% of the ~10,000-Mb haploid Louisiana iris genome. IRRE elements are transcriptionally active in *I. brevicaulis* and *I. fulva* and their $F_1$ and backcross hybrids. The LTRs of the elements are more variable than the coding domains and can be used to define several distinct IRRE subfamilies. Transposon display or S-SAP markers specific to two of these subfamilies have been developed and are highly polymorphic among wild-collected individuals of each species. As IRRE elements are present in each of 11 iris species tested, the marker system has the potential to provide valuable comparative data on the dynamics of retrotransposition in large plant genomes.

THE majority of chromosomal DNA in plants with large genomes is repetitive and is likely composed of various classes of mobile elements (Flavell *et al.* 1974; Joseph *et al.* 1990; Kidwell 2002). Although many classes of elements contribute significantly to overall genome size (*e.g.*, Leeton and Smyth 1993), recent results from the grasses suggest that LTR retrotransposons compose the largest fraction of genomic DNA (SanMiguel *et al.* 1996; SanMiguel and Bennetzen 1998; SanMiguel *et al.* 1998; Bennetzen 2002). In grasses with relatively large genomes such as maize and barley, >60% of the genome is composed of LTR retrotransposons (SanMiguel and Bennetzen 1998; Vicient *et al.* 1999; Meyers *et al.* 2001), while in the smaller rice genome the proportion is 30–35% (N. Jiang, unpublished data).

LTR retrotransposons are class I mobile elements related to infectious retroviruses (Malik *et al.* 2000). There are two major types of LTR retrotransposons, Ty1/*copia*-like (Pseudoviridae) and Ty3/*gypsy*-like (Metaviridae; Hull 1999; Pringle 1999), which are categorized by the order of genes within the *pol* polyprotein (Kumar and Bennetzen 1999). Members of both types are ubiquitous in plant genomes (Flavell *et al.* 1992; Levin 2002). Unlike class II, or DNA elements, which excise from a chromosomal location and insert elsewhere in the genome, class I elements transpose through an RNA intermediate so that a single genomic copy can potentially be the source of numerous new insertions (Kumar and Bennetzen 1999).

Plant retrotransposons have been shown to be activated by several forms of stress to the host plant, including wounding, tissue culture, pathogen attack, and chemical treatment (Grandbastien 1998; Feschotte *et al.* 2002). Wide crosses may also be a source of genomic stress leading to the activation of elements (McClintock 1984). Homoploid interspecific hybridization has been shown to activate LTR retrotransposons in wallabies (Waugh O'Neil *et al.* 1998) and in Drosophila (Labrador *et al.* 1999). However, element activation was not detected in other homoploid interspecific crosses (*e.g.*, Roemer *et al.* 1999; Robinson *et al.* 2000).

The Louisiana iris species complex has a long history as a model system for studying the evolutionary implications of natural hybridization (*e.g.*, Riley 1938; Anderson 1949; Arnold 1997, 2000). The complex consists of four species, *Iris brevicaulis*, *I. fulva*, *I. hexagona*, and the rare hybrid species *I. nelsonii* (Randolph 1966; Arnold 1993). Hybrids involving the first three taxa are common in southeastern Louisiana, especially in areas of recent habitat disturbance (Randolph *et al.* 1967). The work described here is focused on *I. brevicaulis* and *I. fulva*, which are morphologically and karyotypically distinct (Randolph *et al.* 1961), but can produce vigorous hybrids with high fitness (Burke *et al.* 1998).

The goal of this study was to characterize LTR retrotransposons from the large iris genome to take advantage of the abundance and distribution of these elements for the development of molecular markers useful

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY245285–AY245375.

[1]*Corresponding author:* Department of Genetics, Life Sciences Bldg., University of Georgia, Athens, GA 30602.
E-mail: ekentner@arches.uga.edu

for hybridization and speciation research. Two families of related Ty3/*gypsy*-like LTR retrotransposons were characterized using PCR and genomic library screens. These IRRE elements (for *ir*is *r*etro*e*lement) account for 6–10% of the ~9650-Mb iris genome and are transcriptionally active in *I. brevicaulis*, *I. fulva*, and their $F_1$ and backcross hybrids. IRRE elements were detected in each of 11 iris species tested, but not in several related genera. Transposon display or S-SAP primers specific to two subfamilies of IRRE elements were used to generate large numbers of markers in *I. brevicaulis* and *I. fulva*, and the technique can be adapted for use in other iris species as well.

## MATERIALS AND METHODS

**Materials:** All material from Louisiana iris species (*I. brevicaulis*, *I. fulva*, *I hexagona*, and *I. nelsonii*) was obtained from wild-collected plants maintained at the University of Georgia Plant Biology Department greenhouses. Other species were collected from natural populations in Georgia (*I. cristata*, *I. verna*, and *Sisyrinchium* sp.) and California (*I. bracteata*, *I. crysophylla*, *I. douglasiana*, *I. missouriensis*, and *I. longipetala*) or were obtained from plants cultivated in the University of Georgia Plant Biology Department greenhouses (*Acidanthera bicolor* and *Neomarica longifolia*). Seed of the genome size standard *Allium cepa* cv. Ailsa Craig was provided by Michael Bennett (Royal Botanical Garden Kew).

**Nucleic acid extraction:** DNA was extracted using the CTAB procedure of Doyle and Doyle (1987) as modified by Soltis *et al.* (1991) followed by treatment with RNase A. Total RNA was prepared from leaf or root tissue using the RNeasy plant RNA extraction kit (QIAGEN, Valencia, CA) and poly(A)$^+$ RNA was purified from ~600 μg of leaf RNA using an Oligotex mRNA purification kit (QIAGEN). First-strand cDNA was obtained using the Superscript cDNA synthesis kit (Invitrogen, Carlsbad, CA).

**Cloning procedures:** Repetitive elements from *I. fulva* and *I. brevicaulis* were isolated by constructing small insert (~200–900 bp) genomic libraries for each species in the plasmid vector pBlueScript II (Stratagene, La Jolla, CA) following partial digestion of genomic DNA with *Sau*3AI. Libraries were probed with sheared α-$^{32}$P-labeled total genomic DNA (random primers labeling kit, Invitrogen) from either *I. fulva* or *I. brevicaulis*. Plasmid clones showing homology to retrotransposons in database searches were used to probe phage libraries constructed by cloning ~5- to 10-kb *Sau*3AI genomic fragments into the λZAP express phage vector (Stratagene). PCR products were cloned using the TOPO TA cloning kit (Invitrogen).

**Polymerase chain reaction:** Retrotransposon fragments containing the 3′ end of the integrase domain and the 5′ end of the 3′ LTR were amplified using the primer pair LTRSC REENF (CACAYTTGTTYGACTCGTRAGG)/LTRSCREENR (TYRTGCAAGATGTACTTGCC). PCR amplifications were performed on 50–200 ng of genomic DNA in 30-μl reaction volumes containing 1.5 units of Amplitaq DNA polymerase (Perkin Elmer/Applied Biosystems, Foster City, CA), 0.2 mM each dNTP, 1.5 mM $MgCl_2$, and the buffer supplied with the enzyme. Cycling conditions were 94° for 3 min, followed by 32 cycles of 94° for 45 sec, 52° for 45 sec, 72° for 1 min, and ending with 72° for 6 min. Reverse transcriptase (RT)-PCR was performed using the same primers and cycling conditions except that 1 μl of first-strand cDNA or 1 μl of the DNase-treated template RNA for cDNA synthesis (negative controls) was used as a template.

**Transposon display:** Total genomic DNA (~500 ng) was digested overnight at 37° with an excess (50 units) of *Eco*RI. Standard *Eco*RI amplified fragment length polymorphism adapters (Vos *et al.* 1995) were ligated overnight at 25° using 5 units of T4 DNA ligase and the buffer supplied by the manufacturer (Invitrogen). Nested element-specific primers were used in the preamplification and selective amplification reactions with the selective primers closer than the preamplification primers to the element ends.

Preamplification reactions contained 10 pmol of primers homologous to the adapters plus two selective bases (Vos *et al.* 1995) and 10 pmol of primer homologous to the LTR end of either IRRE1-A1 (CCAAACCAAACCAAGCCACACTAA ACC) or IRRE1-C (ACAGGAACACRTTCCAATTACGT). Reactions were performed in 30 μl containing 3 μl of 2:1 diluted restriction/ligation reaction, 1.5 units Ampli*Taq* DNA polymerase (Perkin Elmer/Applied Biosystems), 0.2 mM each dNTP, 2.5 mM $MgCl_2$, and the buffer supplied with the enzyme. The cycling conditions were 72° for 2 min, 94° for 3 min, followed by 30 cycles of 94° for 30 sec, 56° for 30 sec (IRRE1-A1) or 51° for 30 sec (IRRE1-C), 72° for 1 min, and a final elongation of 72° for 3 min.

Selective amplifications were performed in 10 μl containing 1 μl of the 10:1 diluted preamplification reaction, 5 pmol of adapter primer plus four selective bases, 3 pmol $^{33}$P-labeled IRRE1-A1 primer (CGTATAAAATACGTACACAAGAG) or IRRE1-C primer (TCCAATTACGTATAAAATACG), 1.5 units Ampli*Taq* DNA polymerase (Perkin Elmer/Applied Biosystems), 0.2 mM each dNTP, 2.5 mM $MgCl_2$, and the buffer supplied with the enzyme. The cycling conditions were 94° for 3 min, followed by 30 cycles of 94° for 30 sec, 56° for 50 sec (IRRE1-A1) or 51° for 30 sec (IRRE1-C), 72° for 1 min, and a final elongation of 72° for 3 min. The amplification products were run on polyacrylamide sequencing gels and visualized by autoradiography.

**DNA sequencing and analysis:** DNA clones from the plasmid library screens were sequenced by the Molecular Genetics Instrumentation Facility at the University of Georgia. λ-clones and cloned PCR products were sequenced using the Big Dye terminator sequencing kit (Perkin Elmer/Applied Biosystems) on an ABI 377 automated DNA sequencer (Perkin Elmer/Applied Biosystems). A primer walking strategy was employed to sequence the λ-clones and universal sequencing primers were used to sequence the cloned PCR products. DNA and amino acid sequences were aligned with the ClustalW Service at the European Bioinformatics Institute (http://www2.ebi.ac.uk/clustalw) using the default parameters, and GeneDoc (http://www.psc.edu/biomed/genedoc) was used to manually edit and box-shade the alignments. Neighbor-joining trees were constructed using MEGA 2.1 (http://www.megasoftware.net/), and the sliding window analysis was carried out using DnaSp (Rozas and Rozas 1999).

**Flow cytometry:** Nuclear DNA content was measured by flow cytometry according to Galbraith *et al.* (1997). Following the recommendations of Johnston *et al.* (1999), nuclei were prepared simultaneously with those of the plant genome size standard *A. cepa* cv. Ailsa Craig and stained with propidium iodide. Peak fluorescence was measured with a FACSCalibur flow cytometer (BD Biosciences, San Jose, CA) using an excitation wavelength of 488 nm. Cellquest v.3.2.1 (BD Biosciences) was used to analyze the peaks and Flowjo 3.5.4 (Tree Star, San Carlos, CA) was used to generate histograms. Nuclear DNA content was calculated from the peak means according to the formula in Galbraith *et al.* (1997) using $2C = 33.55$ pg as the value for the genome size standard (Johnston *et al.* 1999). All analyses were performed in the CTEGD flow cytometry facility at the University of Georgia.

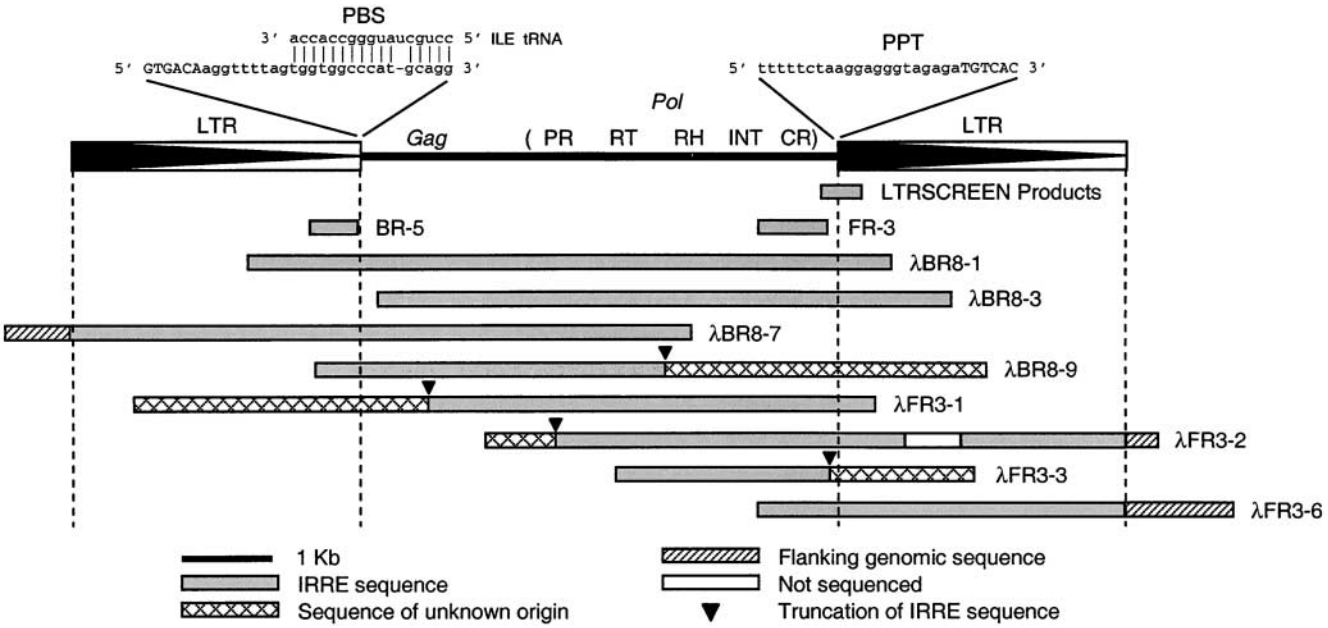


FIGURE 1.—Reconstructed IRRE element based on a series of overlapping λ-clones. LTR are depicted as 3.2 kb in length on the basis of the length of the alignment between the LTRs of FR3-6 (2786 bp) and BR8-7 (3003 bp). The actual range in size of these repeats among IRRE subfamilies is unknown. The sequence of the putative primer binding site (PBS) and its homology to the isoleucine tRNA of *Lupinus luteus* (X06459) is presented at the top of the figure, as are the sequences of the polypurine tract (PPT) and LTR ends (in capital letters). The relative positions of the plasmid copy-number probes (BR-5 and FR3) and the LTRSCREEN PCR products used to define IRRE subfamilies are also depicted. *Pol*, polyprotein; PR, protease; RT, reverse transcriptase; RH, RNAse H; INT, integrase; CR, chromodomain.

**Copy-number determination:** Two methods, dot blot hybridizations and a genomic library screen were used to determine the copy number of the retrotransposon internal domains and LTRs. Two probes labeled with α-$^{32}$P by random priming (Invitrogen) were used. The internal domain probe was FR-3, a 761-bp plasmid clone containing the end of the integrase core domain and the downstream sequence (chromodomain) ending ~100 bp before the start of the 3′ LTR. The LTR probe was BR-5, a 542-bp plasmid clone containing an LTR fragment ending 10 bp before the 3′ end of the LTR (Figure 1).

Serial dilutions of FR-3 or BR-5 and genomic DNA from *I. brevicaulis*, *I. fulva*, *I. hexagona*, and *I. nelsonii* were spotted onto GeneScreen hybridization membranes (New England Nuclear, Boston) using a dot blot apparatus (GIBCO BRL, Gaithersburg, MD). Two replicate dots containing 1, 10, 25, 50, and 100 ng of genomic DNA were made for each species (for a total of 10 dots per species). Internal domain and LTR spots were also replicated twice and contained 0.01, 0.05, 0.125, 0.25, and 0.5 ng of either FR-3 or BR-5. The total amount of DNA in each spot was adjusted to 100 ng with salmon sperm DNA, and the DNA was bound to the membrane using ultraviolet light. DNAs were quantified by fluorimetry (Hoefer Scientific, San Francisco), adjusted to the same concentration, and then checked on agarose gels stained with ethidium bromide before the final dilutions were made. Two identical blots were probed with either FR-3 or BR-5 before a final wash of 0.1× SSC and 0.1% SDS at 65° for 15 min. Hybridization signals from each dot were quantified with a STORM phosphoimager (Molecular Dynamics, Piscataway, NJ) and the average number of counts per copy in the FR-3 and BR-5 dots was used to calculate the total number of copies present in each genomic dot. The genome size measurements obtained by flow cytometry for each species were then used to calculate the number of genomes per dot, and the number of copies of each probe per genome was determined by dividing the number of copies per genomic dot by the number of genomes per dot. As regressions of DNA quantity *vs.* hybridization signal were nearly perfectly linear ($R^2 > 0.99$, data not shown) for each series of dots, the copy numbers reported are the average copy number calculated from all dots of a given species.

Copy number estimates were obtained for FR-3 and BR-5 by screening the *I. brevicaulis* primary λ-phage library (average insert size of ~6900 bp) and counting the number of positive plaques. Replicate filters were made so that the fraction of the library screened was identical for both probes. A total of 3192 plaques containing ~22 Mb were screened and copy numbers were calculated by dividing the number of positive plaques by the proportion of the genome screened (~0.11%).

**Gel blot analysis:** DNA gel blot analysis was performed using GeneScreen hybridization transfer membranes (New England Nuclear) following the manufacturer's "salt transfer protocol" for transferring DNA to the membrane and the "aqueous hybridization buffer for DNA" protocol for prehybridization and hybridization. Following overnight hybridization at 65° the membranes were washed twice with 2× SSC and 1% SDS at 60° for 15 min before a final 15-min wash at 25° with 0.1× SSC.

RNA gel blot analysis was performed as described by SEELEY *et al.* (1992) using ~5 μg of poly(A)$^+$ RNA isolated from *I. brevicaulis* leaf tissue. The blot was subjected to a final wash in 5 mM Tris-HCl pH 8.0 and 0.1% SDS at 65° for 15 min.

## RESULTS

**Isolation and characterization of iris LTR retrotransposons:** The cloning strategy for isolating iris retrotransposons was based on the expectation that the highest copy-number repeats should be LTR retrotransposons.

High copy repetitive sequences were isolated from small-insert *I. brevicaulis* (IB) and *I. fulva* (IF) genomic libraries by probing with sheared total DNA from the genome used to construct the library. Sixteen IB clones and 12 IF clones were recovered and 7 randomly chosen clones from each species were confirmed to be repetitive by DNA gel blot hybridization (data not shown) before all 28 clones were fully sequenced. BLASTX searches revealed that 10 of the 28 clones share sequence similarity with the coding regions of LTR retrotransposons in the public databases.

To obtain the LTR sequence information necessary for the development of the primers for transposon display (see below), λ-phage libraries were constructed and probed. Two clones, FR3 (*fulva* repeat 3) and BR8 (*brevicaulis* repeat 8), were chosen to probe *I. fulva* and *I. brevicaulis* λ-phage libraries, respectively, on the basis of their high level of amino acid similarity to Ty3/*Gypsy*-like elements in the databases. FR3 is a 761-bp integrase/chromodomain fragment and BR8 is a 426-bp RNaseH fragment. Both probes hybridized strongly with >5% of the plaques screened. Six *I. fulva* clones hybridizing to the FR3 probe (λFR3s) and eight *I. brevicaulis* clones hybridizing to the BR8 probe (λBR8s) were chosen for DNA sequencing. Three of the λFR3 clones and four of the λBR8 clones were fully sequenced and the rest of the clones were partially sequenced from each end. The sequencing of a clone was abandoned when it became clear that it did not contain fragments useful for defining the LTR ends (our primary objective) or, in a few cases, when regions that were difficult to sequence were encountered.

The sequence of the larger fragments contained in the λ-clones revealed that both of the probes were fragments of elements belonging to closely related Ty3/*Gypsy*-like retrotransposons. The elements were named IRRE, using the naming scheme that has been applied to rice (RIRE; Nakajima *et al.* 1996), barley (BARE; Manninen and Schulman 1993), oat (OARE; Kimura *et al.* 2001), and other LTR retrotransposons. The elements contain genes arranged in the order typical of Ty3/*Gypsy*-like elements (Kumar and Bennetzen 1999) and contain a putative chromatin binding domain (Malik and Eickbush 1999) downstream of the integrase gene (Figure 2). Consistent with this observation, IRRE elements group with other plant chromodomain-containing LTR retrotransposons in a neighbor-joining tree based on an amino acid alignment of the RT domain (Figure 3). The phylogenetic analysis also revealed two well-supported groups of iris elements (>90% of bootstrap replications) that were named IRRE1 and IRRE2 (Figure 3). Following the recommendation of Bowen and McDonald (1999), these two groups of elements are referred to as distinct "families" because they display >10% divergence in the amino acid sequence of their RT domains. The two families are also clearly differentiated by amino acid substitutions in the additional protein core domains presented in Figure 2.

While retrotransposon proteins are well conserved and easily recognizable, LTR sequences are highly variable in length and in primary sequence and generally cannot be identified for uncharacterized elements using database searches. Instead, LTRs must be defined as direct repeats flanking the coding region of an element. Attempts to define the IRRE LTRs using this strategy were complicated by the length of the LTRs relative to the average insert size of the libraries from which the clones were derived (λ-phage library average insert sizes: *I. fulva*, ~6200 bp; *I. brevicaulis*, ~6900 bp), so a complete IRRE sequence was reconstructed from a series of overlapping λ-clones representing paralogous copies of the element (Figure 1). Variable, but identifiable direct repeats of ~2.8–3.0 kb flanking the coding region of several clones were identified as likely LTRs. The LTRs end in the typical 5′ TG preceded by a polypurine tract (PPT) and in a 3′ CA followed by a primer binding site (PBS). The putative PBS is most similar to the cytoplasmic isoleucine tRNA from *L. luteus* (Figure 1; Barciszewska *et al.* 1988), which is unusual, as the PBS of most (but not all) plant retrotransposons is derived from a methionine tRNA (Kumar and Bennetzen 1999). LTRs typically end in short inverted repeats, and the putative IRRE LTRs end in the 6-bp inverted repeat 5′-TGTCAC/GTGACA-3′. For additional confirmation that the LTR ends had been properly defined, the sequences flanking the putative LTR ends were compared among all of the clones containing these sequences (13 clones, both plasmid and λ). In all cases, the sequence similarity between the clones either dropped off abruptly at the end of the LTRs (representing the flanking genomic DNA) or continued into the coding regions (either the *gag* or the integrase) of the element, as expected (data not shown).

**The iris genome contains diverse subfamilies of IRRE elements:** Alignment of the LTR sequences from the λ-clones clearly indicated that the IRRE1 and IRRE2 families can be divided into subfamilies of elements sharing diagnostic nucleotide residues at many positions. To further define these subfamilies and to derive the LTR-end consensus sequences necessary for the design of transposon display primers, the PCR primer pair LTRSCREENF/LTRSCREENR was used to amplify IRRE fragments consisting of the noncoding region after the stop codon of the *pol* domain and the first ~280 bp of the 3′ LTR (Figure 1). These primers are degenerate and were designed to amplify as many IRRE variants as possible given the available sequence information. A total of 34 of these PCR products were cloned from genomes of *I. brevicaulis* and *I. fulva* and sequenced, revealing remarkable LTR diversity among IRRE elements. The relationships among IRRE subfamilies as defined by these LTR sequences is presented in the neighbor-joining tree of Figure 4. While the adja-

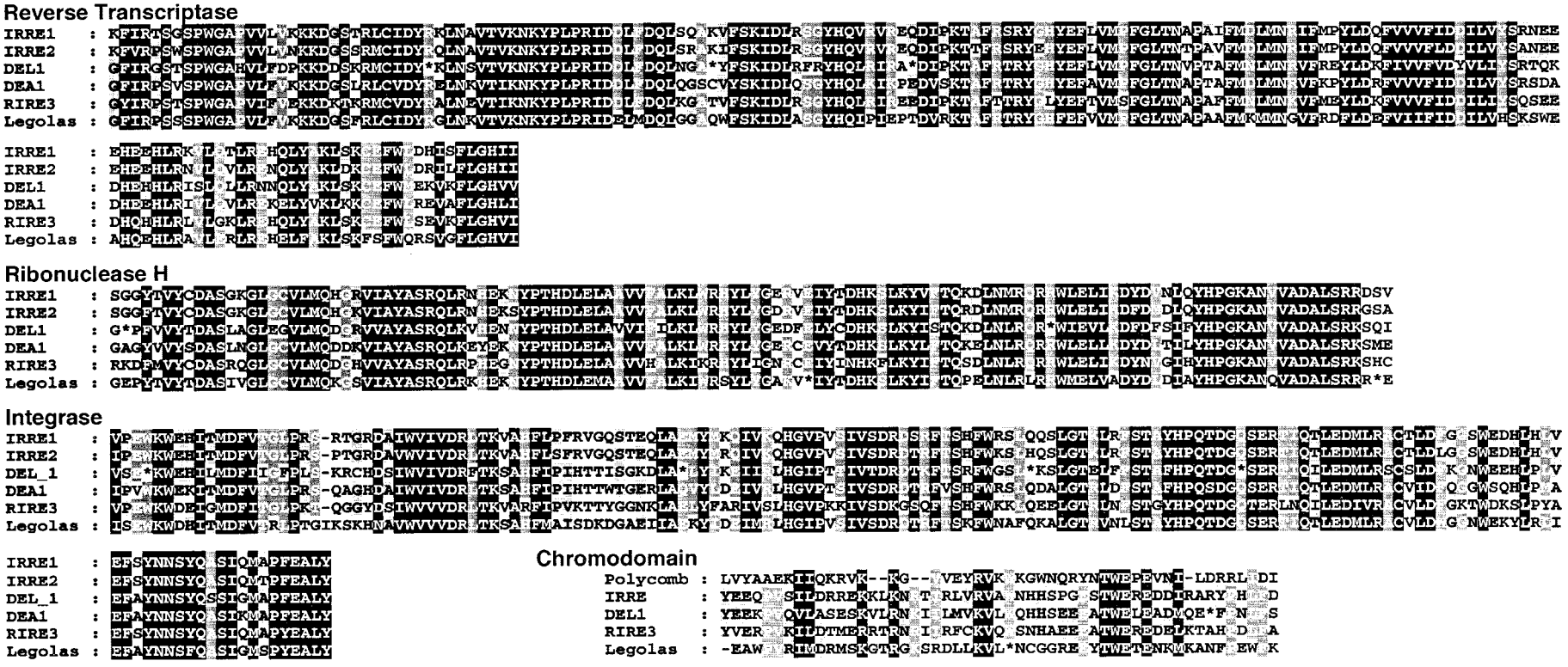


FIGURE 2.—Amino acid alignments of protein domains from IRRE and other chromodomain-containing plant retrotransposons. The IRRE1 and IRRE2 sequences presented are consensus sequences. There were insufficient data to reconstruct the IRRE2 consensus for the chromodomain, and the sequence presented is for IRRE1. The alignment of each protein domain consists of the residues present in the "core domain" defined by the Pfam database (http://pfam.wustl.edu/). Other LTR retrotransposons are: *del1*, *Lilium henryi* (X13886); *dea1*, *Ananas comosas* (Y12432); *RIRE3*, *Oryza sativa* (AB014738); and *Legolas*, *Arabidopsis thaliana* (AC007730.1).

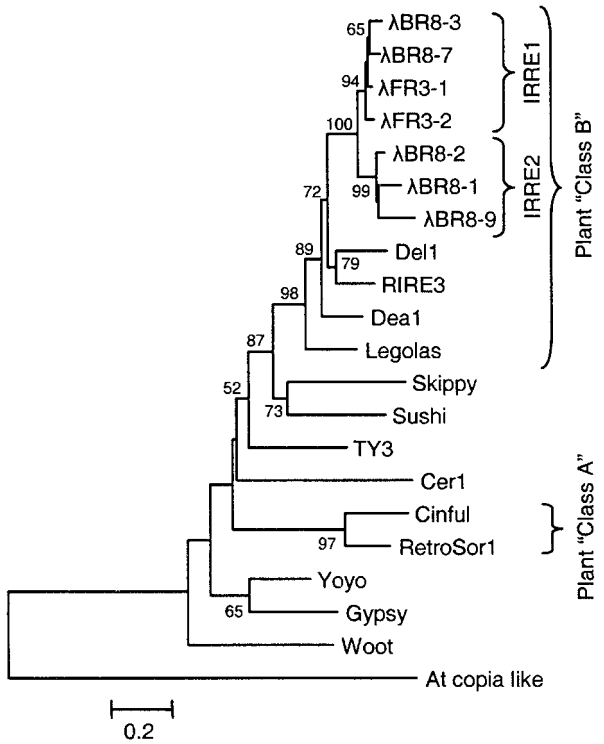


FIGURE 3.—Phylogenetic relationships of Louisiana iris retrotransposons to other Ty3/*gypsy*-like retrotransposons. The tree is based on an amino acid alignment of the reverse transcriptase domain using the neighbor-joining algorithm and was rooted using a *copia*-like element from *A. thaliana* (AAG-51258). Branches receiving <50% support in 1000 bootstrap replications are not shown. Plant "Class A" and "Class B" groups after MARIN and LLORENS (2000) correspond to previously described clades of Ty3/*gypsy*-like elements. The scale bar depicts Poisson-corrected distances. Elements and the organisms from which they were isolated are: *del1*, *L. henryi* (X13886); *dea1*, *A. comosas* (Y12432); *RIRE3*, *O. sativa* (AB-014738); *Legolas*, *A. thaliana* (AC007730.1); *Skippy*, *Fusarium oxysporum* (AAA88791); *Sushi*, *Fugu rubripes* (AAC335260); *Cer1*, *Caenorhabditis elegans* (AAA50456); *TY3*, *Saccharomyces cerevisiae* (S69842); *Cinful*, *Zea mays* (T14595); *RetroSor1*, *Sorghum bicolor* (AAD19359); *Yoyo*, *Ceratitis capitata* (AAC28743); *Gypsy*, *Drosophila melanogaster* (GNFFG1); and *Woot*, *Tribolium castaneum* (AAC47271).

cent internal domain is relatively conserved among all of the sequenced PCR products, the LTR sequences contain numerous insertion/deletion polymorphisms of ~3–30 bp that are most often shared by several sequences. The overall size of the region corresponding to these PCR products among 59 genomic and cDNA sequences (see below) varies from 382 to 498 bp.

Retrotransposons are transcribed from promoter elements typically located in the 5′ end of the LTR. To identify potential IRRE promoter sequences, each PCR product was analyzed with eukaryotic promoter prediction software (http://www.fruitfly.org/seq_tools/promoter.html). A region of the LTR was consistently identified as a likely (score ≥0.90) TATA box and transcriptional start site. To confirm this result, a sliding window analysis of nucleotide diversity across the alignment of genomic PCR products was performed to search for conserved and, therefore, possible functional domains within the IRRE LTR sequences (Figure 5). Two highly conserved regions were detected, one of them corresponding to the putative PPT/LTR end and the other located ~150 bp downstream in the alignment. This second conserved region corresponds to the putative promoter sequences independently identified by the promoter prediction software.

**Louisiana iris genome size:** Because an estimation of total genome size is required to calculate the copy number of IRRE elements, flow cytometry was used to measure the C values of each of the four hybridizing Louisiana iris species (Figure 6). The values measured for each species (*I. brevicaulis*, $2C = 19.75$ pg; *I. fulva*, $2C = 19.57$ pg; *I. hexagona*, $2C = 19.59$ pg; *I. nelsonii*, $2C = 20.04$ pg) are comparable to the available data for other Iris species (iris median is 19.05 pg; BENNETT *et al.* 1998). All of the Louisiana iris species appear to have similarly sized genomes, but valid comparisons at the species level are not possible because only a single individual of each was measured. The size of these genomes is large relative to other angiosperms, as *I. fulva* has a larger genome than ~81% of the ~3400 species that have been measured (BENNETT *et al.* 1998).

**Copy-number estimation of IRRE elements:** Two methods were used to determine the genomic copy number of IRRE elements, dot blot hybridizations and a genomic library screen. Estimates were obtained for each of the four hybridizing Louisiana iris species using dot blots and independently estimated for *I. brevicaulis* by screening the primary phage library used to isolate the IRRE clones (the *I. fulva* library was amplified and was therefore inappropriate for copy-number determination). The results obtained for both methods are presented in Table 1 and indicate that between $6.5 \times 10^4$ and $1 \times 10^5$ copies of IRRE elements are present per haploid genome. Assuming an average element size of 11 kb, IRRE sequences are estimated to account for ~6–10% of the Louisiana iris genome.

**Estimation of the number of solo LTRs:** Recombination between the LTRs of a retrotransposon can result in the loss of internal sequences, leaving behind a solo LTR. The ratio of intact elements to solo LTRs for the barley retrotransposon *BARE-1* has been shown to be highly variable among barley species, with the excess of LTR sequences reported to be 7- to 42-fold greater than the expected two-to-one ratio (VICIENT *et al.* 1999). In an effort to determine the ratio of intact IRRE elements to solo LTRs in the Louisiana iris, the copy number of both an internal and an LTR probe were calculated (Table 1). However, in all cases, fewer LTR sequences than the expected minimum ratio of two to one were detected on the basis of the calculated number of internal regions (Table 1). This result is most likely due to the rapid divergence of the noncoding LTR sequences relative to the more highly conserved element coding domains as the LTR sequences of the λ-clones differ by

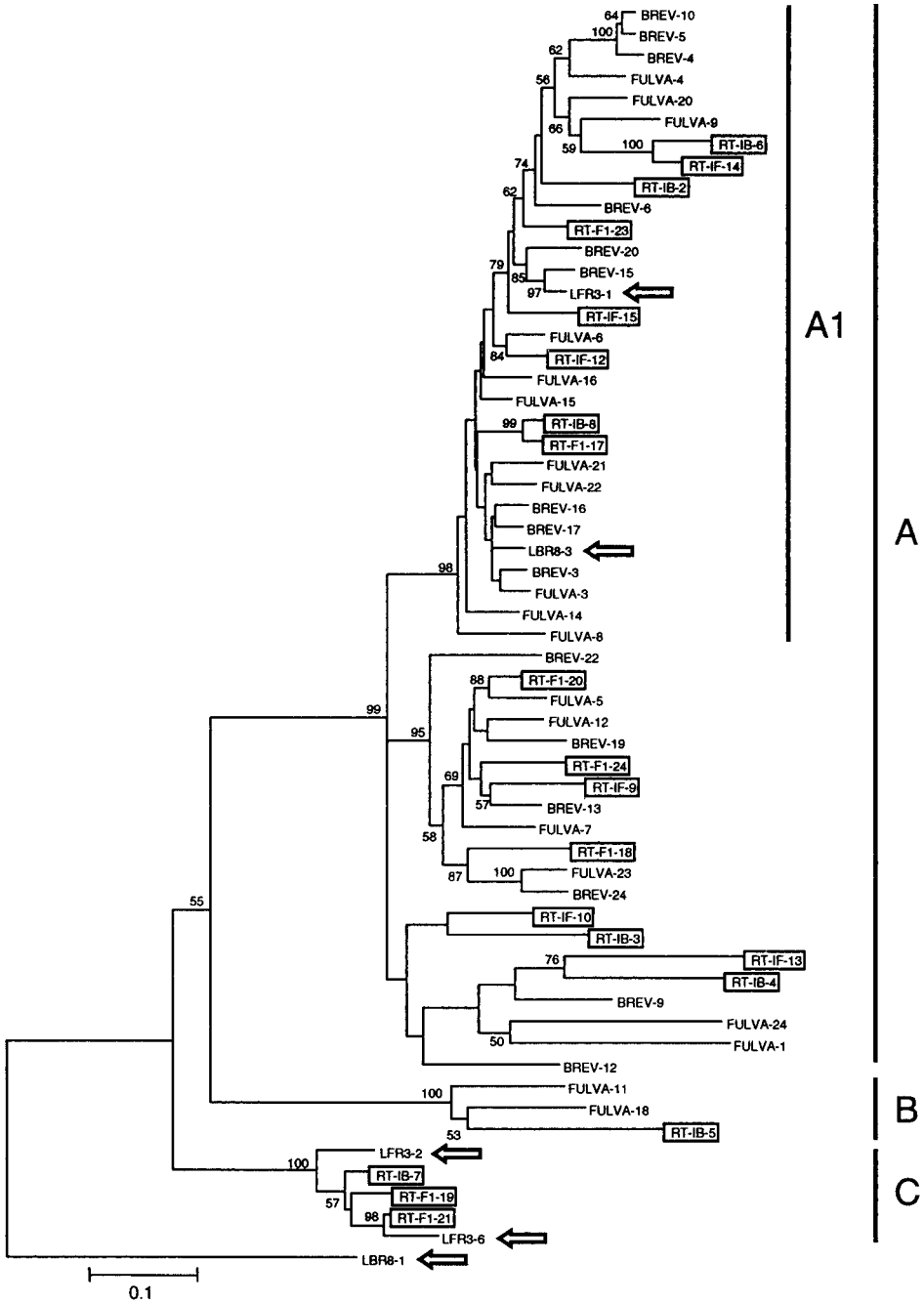


FIGURE 4.—Neighbor-joining tree of IRRE LTR ends and adjacent internal region. The tree is based on the alignment of RT-PCR and genomic PCR products from the primer pair LTRSCREENF/LTRCREENR and of λ-clones containing the same region, which includes ~200 bp of the internal domain downstream of the *pol* stop codon and the first ~280 bp of the 3′ LTR. RT-PCR products are boxed and arrows indicate the λ-clones. Numbers on the branches represent the percentage of bootstrap support calculated from 1000 replicates, with values <50% not shown. The tree is rooted with the IRRE2 element λBR8-1. The scale bar depicts distances on the basis of the Kimura two-parameter substitution model. Bars and letters (A, A1, B, and C) to the right of the tree indicate IRRE1 subfamily designations.

as much as 30% in the probe region and the hybridization wash conditions were stringent. If the LTRs evolve faster than the internal domains, then the LTR probe would hybridize to fewer IRRE subfamilies than the internal probe would, resulting in an underestimate of the number of LTRs. To test for this possibility, comparisons of nucleotide similarity for the region homologous to the LTR probe (BR5) and for the element protein core domains were made among all pairs of λ-clones containing the appropriate sequences. In these comparisons, the nucleotide sequence of the region homologous to the LTR probe is significantly more divergent among element copies than are the coding regions (randomization test, $P < 0.001$), suggesting that the region of the IRRE LTR corresponding to the probe evolves at a faster rate than the element coding regions.

**IRRE elements are transcriptionally active:** To test for the possible transcriptional activity of IRRE elements, an *I. fulva* × *I. brevicaulis* interspecific mapping population was assayed using RT-PCR. The parents (*i.e.*, "pure" *I. brevicaulis* and *I. fulva*), several $F_1$ plants, and five back crosses to each parent were assayed for IRRE transcripts using the LTRSCREEENF/LTRSCREENR primer pair. Transcripts were present in all of the genotypes tested (Figure 7). Contamination by genomic DNA was ruled by negative controls, which used the DNase-treated RNA as the amplification template. For *I. brevicaulis*, transcripts were also detected on Northern blots (data not

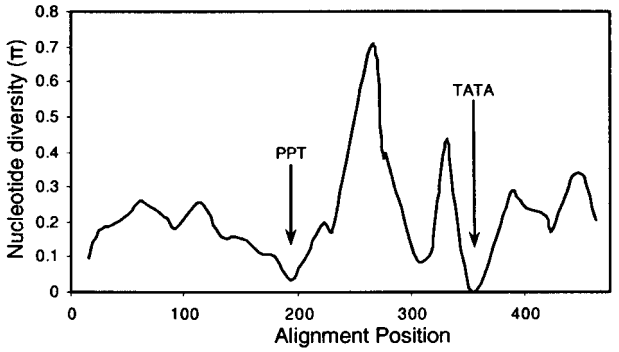


FIGURE 5.—Nucleotide diversity across an alignment of 34 LTR sequences and adjacent noncoding internal domain. Sliding window analysis of genomic LTRSCREEN products show regions of low diversity corresponding to the PPT/LTR start (PPT) and to the predicted promoter sequence (TATA). A 25-bp window length and step size of 12 bp was used with the pairwise gap deletion option of DNAsp selected.

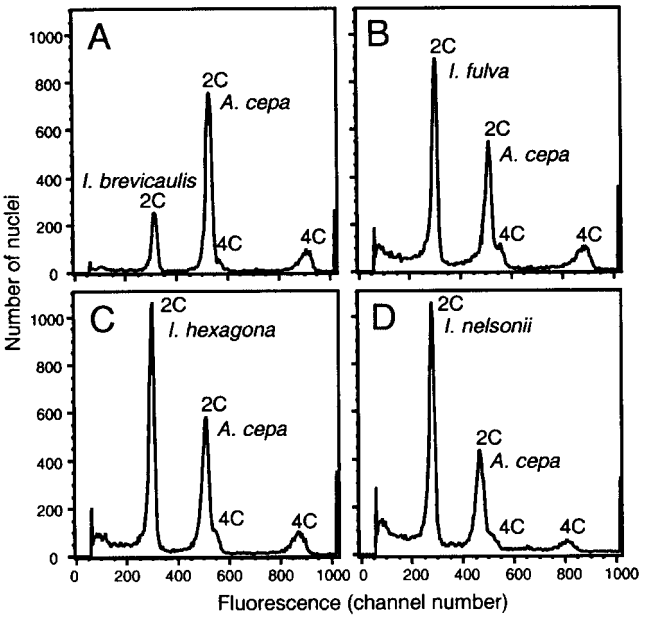


FIGURE 6.—Flow histograms of nuclear DNA content in the Louisiana iris with *A. cepa* cv. Ailsa Craig nuclei used as the genome size standard. Coefficients of variation for Iris and Allium 2*C* peaks, respectively, are (A) 3.28%, 2.25%; (B) 3.49%, 2.58%; (C) 3.40%, 2.68%; (D) 3.91%, 2.97%.

shown), but only when a relatively large amount of poly(A)$^+$ RNA (~5 μg) was used, suggesting that IRRE transcripts are not particularly abundant. To verify that the amplified bands represent IRRE fragments, 21 cloned PCR products were sequenced from *I. brevicaulis*, *I. fulva*, and an $F_1$ hybrid between them. The two bands evident in all of the RT-PCR reactions represent different subfamilies of elements containing insertion/deletion polymorphisms with the larger band representing at least two sequence variants that result in similar overall fragment length.

**IRRE retrotransposons are useful molecular markers:** One of the primary reasons for characterizing LTR retrotransposons from Louisiana iris species was to develop transposon display or S-SAP markers (WAUGH *et al.* 1997; VAN DEN BROECK *et al.* 1998). This marker technology is attractive for use in the Louisiana iris system because it takes advantage of the abundant repeats that are characteristic of plants with large genomes. To this end, PCR primers were developed on the basis of the consensus sequence of the ends of the LTRs for two IRRE1 subfamilies, IRRE1-A1 and IRRE1-C (Figure 4). The subfamily designated IRRE1-A1 contains both genomic and RT-PCR products, whereas the IRRE1-C subfamily contains only RT-PCR products and two of the λ-clones. The LTR sequence of the two subfamilies is divergent in the region suitable for transposon display primer sites, enabling the design of subfamily-specific primers.

To test the level of polymorphism of the IRRE retrotransposon-based markers, 10 wild-collected individuals each of *I. fulva* and *I. brevicaulis* were screened using primers specific for the IRRE1-A1 and IRRE1-C subfamilies (Figure 8). Both sets of primers amplified numerous bands from each species, and a high proportion of these bands are polymorphic among the individuals tested (Table 2). Several of the monomorphic bands appear to be species-specific markers (Table 2), which are particularly useful for studying natural hybridization. However, transposon display generates dominant markers, and high-frequency insertions may not be distinguishable from fixed insertions when the number of individuals sampled is small. Assuming that the monomorphic bands in the sample are fixed, the proportion of polymorphic loci is significantly different between the two species for both elements (exact test: IRRE1-A1, $P < 0.0001$; IRRE1-C, $P = 0.023$). No significant difference in the level of polymorphism between the two element subfamilies was detected within either species (exact test: *I. brevicaulis*, $P = 0.813$; *I. fulva*, $P = 0.085$), suggesting that the timing and/or level of retrotranspositional activity is not dramatically different for the two subfamilies. The majority of bands generated for both subfamilies is likely to represent individual loci as they segregate in normal Mendelian ratios in a separate set of linkage mapping experiments using these markers (A. BOUCK, E. KENTNER, R. PEELER, M. ARNOLD and S. WESSLER, unpublished data).

**IRRE retrotransposons are present in many iris species:** To investigate the taxonomic distribution of IRRE LTR retrotransposons, we assayed their presence in 11 iris species and in three other genera of Iridaceae by PCR and/or Southern hybridizations (Figure 9). The results for both techniques were consistent in all cases. IRRE elements are present in all members of the genus Iris examined, although the hybridization signal on Southern blots is much stronger in the Louisiana iris than in other members of the genus (Figure 9). This result could be due to the sequence divergence of IRRE

**TABLE 1**

**IRRE copy number**

| Iris species | INT | LTR | LTR/INT |
|---|---|---|---|
| *I. brevicaulis*[a] | 204,048 | 374,530 | 1.84 |
| *I. brevicaulis*[b] | 128,355 ± 9,777 | 180,098 ± 27,937 | 1.40 |
| *I. fulva*[b] | 188,459 ± 14,652 | 253,384 ± 43,900 | 1.34 |
| *I. hexagona*[b] | 163,843 ± 22,451 | 198,612 ± 46,202 | 1.21 |
| *I. nelsonii*[b] | 154,752 ± 28,582 | 209,642 ± 45,701 | 1.35 |

[a] Genomic library screen.
[b] Dot blot. The mean ±SD of 10 replicate dots is reported.

elements in the genomes of more distantly related iris, lower IRRE copy number in these genomes, or both. For the California irises (Figure 9A, lanes 5–7) preliminary results obtained by sequencing IRRE PCR products suggest that the lower hybridization signal may be due to sequence divergence (E. Kentner, unpublished data).

## DISCUSSION

The IRRE elements are typical Ty3/*Gypsy*-like LTR retrotransposons that occur in high copy number in the genomes of each of the four species of hybridizing Louisiana iris. LTR retrotransposons are major components of plant genomes, and the phylogenetic relationships among a diverse set of these elements or element fragments from many plant species have been determined (*e.g.*, Marin and Llorens 2000). To determine the evolutionary placement of the IRRE elements, we aligned IRRE RT sequences with those of representatives from each major clade of the existing phylogenies and generated a neighbor-joining tree. The group of elements to which the IRRE elements belong was originally identified by Wright and Voytas (1998; Plant Branch 1), with additional elements assigned to it by Marin and Llorens (2000; Plant Class B). The members of this group are characterized by having a putative chromatin-binding domain downstream of the integrase gene. The clade seems to be ancient and ubiquitous in plants since it includes elements from monocots, dicots, and gymnosperms. As IRRE elements are most closely related to elements from other monocots (Figure 3), it appears that the primary mode of transmission for these elements has been vertical.

Several subfamilies of IRRE elements can be distinguished on the basis of the sequence variation in their LTR ends. This variation is similar to the variation documented among the Tnt1 subfamilies of tobacco (Casacuberta *et al.* 1995, 1997; Vernhettes *et al.* 1998) and among copies of the Retrolyc1 retrotransposon of Lycopersicon (Araujo *et al.* 2001), for which the promoter region is variable among subfamilies while the adjacent internal region is more highly conserved. For Tnt1, this promoter variation is correlated with the expression of specific subfamilies in response to different stress-associated signaling molecules (Beguiristain *et al.* 2001), suggesting that adaptive promoter variants have arisen through the error-prone process of retrotransposition (Preston 1996; Casacuberta *et al.* 1997; Beguiristain *et al.* 2001). Although this study has identified putative promoter elements within the IRRE LTRs, it is currently unknown whether the variation present in these sequences has an influence on the replication cycle of the retrotransposons. From a practical standpoint, it is fortunate that the most highly variable LTR region among IRRE subfamilies (Figure 5) corresponds to the optimal region for transposon display primers.

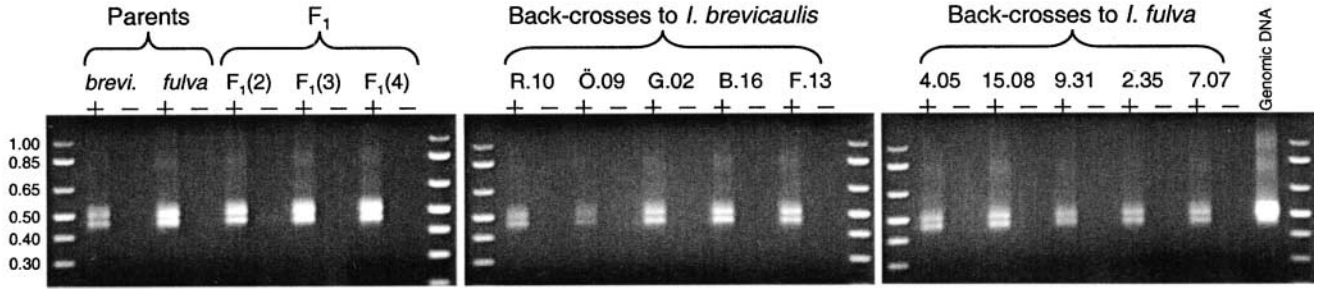


Figure 7.—RT-PCR amplification of iris retrotransposon sequences. +, lanes containing cDNA as the PCR template; −, lanes containing RNA untreated with reverse transcriptase as a control for contamination by genomic DNA. Lane designations in the backcross panels refer to individual genotypes from the mapping population. Products include the end of the putative chromatin-binding domain and the first ~280 bp of the LTR. Double bands are the results of insertion-deletion polymorphisms among LTR sequences.

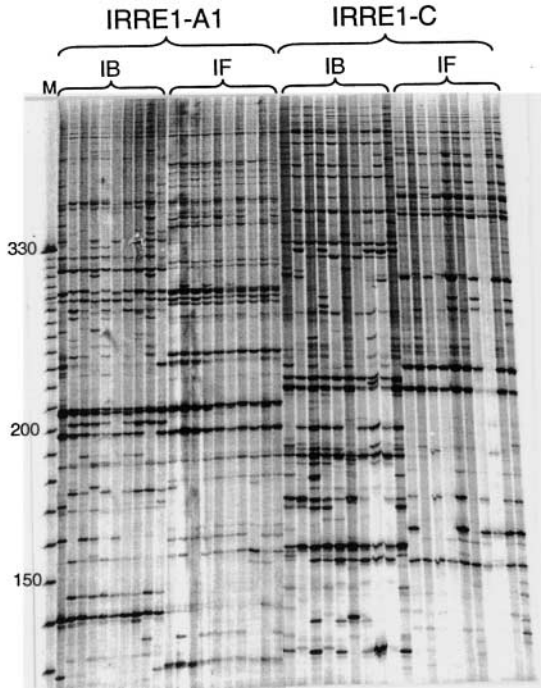


FIGURE 8.—Transposon display using primers specific to the two element subfamilies IRRE1-A1 and IRRE1-C. Both sets of reactions used the four selective bases CTAT. Ten wild-collected individuals of each species were used. IB, *I. brevicaulis*; IF, *I. fulva*. Size markers to the left are in number of base pairs.

This has facilitated the development of subfamily-specific primers that will make comparative studies of polymorphism among element subfamilies possible.

Although the 3′ ends of the IRRE LTRs are less variable than the 5′ ends containing the putative promoter elements, the sequence of the 3′ end of the LTR corresponding to the copy number probe is more variable among IRRE copies than is the sequence of the internal probe. Given the level of LTR variation among IRRE subfamilies and the size of the iris genome, the accurate quantification of the ratio of intact elements to solo LTRs may require an alternative strategy such as the construction and screening of BAC libraries, which would be very difficult considering the size of the iris genome. As discussed by MEYERS *et al.* (2001), there are limitations to measuring copy number with hybridization-based techniques because both sequence divergence among repetitive elements and copy number can affect the hybridization signal. Also, the accuracy of the dot blot technique for determining copy number is dependent on the quantification of the DNA in the dots unless a probe for a single-copy gene is used as an internal control. The copy-number estimates from the library screening do not depend on DNA quantification and may be more accurate. However, both methods indicate that at least 800 Mb of the haploid iris genome is composed of IRRE elements.

At $\sim$0.75–1.0 $\times$ $10^5$ copies, IRRE elements are abundant in the Louisiana iris genome, but well within the range that has been observed for LTR retrotransposons in other plant genomes. For example, the Ty3/*gypsy*-like element Huck accounts for $\sim$10% of the 2.5 $\times$ $10^9$-bp maize genome with a copy number exceeding 1 $\times$ $10^5$ (MEYERS *et al.* 2001). An element closely related to IRRE, *del*1, is present in 1.3 $\times$ $10^4$ copies in *L. henryi*, but the copy number is variable among Lilium species and is not correlated with the relationships between species (JOSEPH *et al.* 1990). IRRE copy number also appears to be variable among iris species (Figure 9), although sequence divergence may also contribute to the hybridization pattern. The copy number of *BARE-1* in wild barley is correlated with microclimatic conditions and varies more than threefold among individuals within a single canyon (KALENDAR *et al.* 2000). It will be interesting to compare the BARE-1 results to the situation in iris by investigating the insertional polymorphism of IRRE elements in natural iris populations.

To date, only a handful of plant LTR retrotransposons have been shown to be transcriptionally active, with activation most often associated with biotic or abiotic stresses (GRANDBASTIEN 1998; FESCHOTTE *et al.* 2002). The exceptions seem to be the *BARE-1* element from barley and the related *OARE-1* from oat for which low levels of transcription are detectable under normal growing conditions (SUONIEMI *et al.* 1996; KIMURA *et al.* 2001), although *OARE-1* transcription is also upregulated by stress (KIMURA *et al.* 2001). Like these elements, the high-copy IRRE elements are expressed under normal growing conditions (Figure 7). However, retrotransposition can be controlled post-transcriptionally, and transcribed elements do not necessarily produce new insertions (CURCIO and GARFINKEL 1999). The maize genome appears to have reached its present size

**TABLE 2**

**Polymorphism detected by IRRE transposon display**

| Element subfamily | Species | Total no. of bands | Polymorphic loci (%) | No. of species-specific bands |
|---|---|---|---|---|
| IRRE1-A1 | *I. brevicaulis* | 63 | 84.1 | 2 |
| | *I. fulva* | 53 | 45.3 | 3 |
| IRRE1-C | *I. brevicaulis* | 61 | 82 | 6 |
| | *I. fulva* | 56 | 62.5 | 8 |

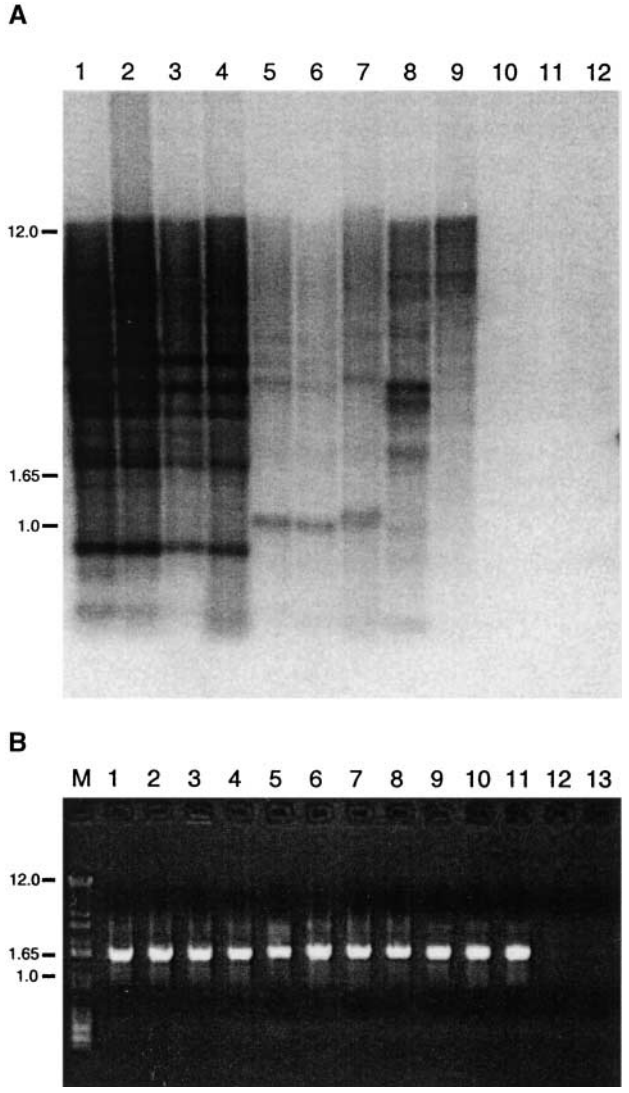


Figure 9.—Survey of IRRE-like retrotransposons in the genus Iris and in other Iridaceae. (A) Genomic DNA gel blot hybridization of the integrase/chromodomain probe FR3 to the genomic DNAs of 12 species of Iridaceae. All lanes contain ~2 μg of genomic DNA digested with *Eco*RI: *I. brevicaulis* (lane 1), *I. fulva* (lane 2), *I. hexagona* (lane 3), *I. nelsonii* (lane 4), *I. bracteata* (lane 5), *I. douglassiana* (lane 6), *I. inominata* (lane 7), *I. cristata* (lane 8), *I. verna* (lane 9), *A. bicolor* (lane 10), *N. longifolia* (lane 11), and *Sisyrinchium* sp. (lane 12). (B) PCR amplification IRRE fragments (integrase core domain plus ~280 bp of the 3′ LTR) from iris species. *I. brevicaulis* (lane 1), *I. fulva* (lane 2), *I. nelsonii* (lane 3), *I. hexagona* (lane 4), *I. verna* (lane 5), *I. cristata* (lane 6), *I. missouriensis* (lane 7), *I. longipetala* (lane 8), *I. douglassiana* (lane 9), *I. bracteata* (lane 10), *I. inominata* (lane 11), *Sisyrinchium* sp. (lane 12), and *Neomarica longifolia* (lane 13).

through recent bursts of retrotransposon activity (San-Miguel and Bennetzen 1998). An interesting but unresolved question is whether the iris genome has reached its present size through such bursts of retrotransposition or through continuous element activity as suggested by the transcription data.

McClintock (1984) predicted that interspecific hybridization may be a form of genomic stress that could lead to the mobilization of transposable elements. Indeed, interspecific hybridization in wallabies is associated with genome-wide loss of DNA methylation and a massive amplification of retrotransposons within a single generation (Waugh O'Neil *et al.* 1998). A less dramatic, but significant, increase in retrotransposition has also been documented following interspecific hybridization in Drosophila (Labrador *et al.* 1999). To investigate the possibility that hybridization between Louisiana iris species could lead to the transcriptional activation of IRRE retrotransposons, a backcross interspecific mapping population was assayed for element-encoded transcripts by RT-PCR. However, transcripts were present in all of the hybrid and pure species individuals tested, and there was no evidence that previously quiescent IRRE elements were activated following hybridization. While the RT-PCR results seem to rule out the kind of retrotransposition burst observed in wallaby hybrids, no conclusion can be reached regarding more subtle changes in the level of transcription in the various hybrids with the current data. If retrotansposition is occurring in the hybrids, new insertions may be very difficult to detect given the number of existing IRRE elements in the iris genome. To date, no genetic evidence for new insertions has been observed in a large sample of backcross hybrids in a mapping population that has been genotyped extensively with IRRE transposon display markers (A. Bouck, E. Kentner, R. Peeler, M. Arnold and S. Wessler, unpublished data).

Transposon display markers were developed for two subfamilies of IRRE elements and insertional polymorphism was assayed in wild-collected individuals of *I. brevicaulis* and *I. fulva*. For markers derived from both subfamilies of elements, the proportion of polymorphic loci is higher for *I. brevicaulis* than for *I. fulva*. The allozyme data of Arnold *et al.* (1990) show the same trend in species-level polymorphism with *I. brevicaulis* containing a higher proportion of polymorphic loci (54%) than of *I. fulva* (45%). The standing level of polymorphism detected by any marker system can be influenced by many aspects of a species' population biology (*e.g.*, Hamrick and Godt 1996; Charlesworth and Wright 2001), but this fact has often been ignored in the literature in favor of arguments equating the insertional polymorphism of transposons with recent element activity. That the timing of insertion events cannot necessarily be inferred from the existence of polymorphism has been clearly demonstrated in maize, where sequencing data have shown that polymorphisms generated by a burst of retrotransposition estimated to have occurred 2–3 MYA are still segregating in modern North American maize lines (Fu and Dooner 2002). Currently, very little data exist on the population genetics of plant retrotransposons. If, in contrast to the situation in Drosophila where most euchromatic retrotransposon insertions are likely to be deleterious (Charlesworth and Langley 1989; Bartolome *et al.* 2002; Carr *et al.* 2002), the average IRRE insertion is

neutral with respect to plant fitness, then the polymorphism of IRRE insertions is likely to be influenced by the population biology of the iris species, as suggested by the allozyme data. Although no data currently exist pertaining to the fitness effects of IRRE insertions, it is difficult to imagine these elements attaining such a high copy number if new insertions are most often deleterious.

Retrotransposons closely related to the IRRE elements cloned from *I. brevicaulis* and *I. fulva* are present in each of 11 iris species tested. The sample includes a representation of species belonging to the subgenus Limniris (the beardless iris), and it is likely that all native North American iris contain these elements. The LTR ends of IRRE elements can be readily amplified from all of these species using the degenerate primers, and these products can be cloned and sequenced using standard techniques. The sequence of the LTR ends can then be used to define additional IRRE subfamilies for transposon display development. As the preliminary sequencing of LTR ends from several species outside of the series Hexagonae (the Louisiana iris) have yielded divergent complements of IRRE subfamilies (E. Kentner, unpublished data), the application of these markers to other iris species may require this additional step of subfamily discovery and definition. The markers should be useful for many applications in evolutionary biology and genetics and it will be interesting to compare insertional polymorphism among IRRE subfamilies and among iris species to gain insight into the dynamics of retrotransposition in large plant genomes.

We thank Mike Scanlon, Brad Bernstein, and Ryan Peeler for assistance in the laboratory and Ning Jiang, Cedric Feschotte, Scott Cornman, and Amy Bouck for critical reading of the manuscript. E.K.K. was supported by a training grant from the National Science Foundation (DBI 9602223).

## LITERATURE CITED

Anderson, E. A., 1949 *Introgressive Hybridization.* John Wiley & Sons, New York.

Araujo, P. G., J. M. Casacuberta, A. P. P. Costa, R. Y. Hashimoto, M. A. Grandbastien *et al.*, 2001 Retrolyc1 subfamilies defined by different U3 LTR regulatory regions in the Lycopersicon genus. Mol. Genet. Genomics **266:** 35–41.

Arnold, M. L., 1993 *Iris nelsonii* (Iridaceae): origin and genetic composition of a homoploid hybrid species. Am. J. Bot. **80:** 577–583.

Arnold, M. L., 1997 *Natural Hybridization and Evolution.* Oxford University Press, Oxford.

Arnold, M. L., 2000 Anderson's paradigm: Louisiana irises and the study of evolutionary phenomena. Mol. Ecol. **9:** 1687–1698.

Arnold, M. L., J. L. Hamrick and B. D. Bennett, 1990 Allozyme variation in Louisiana irises: a test for introgression and hybrid speciation. Heredity **65:** 297–306.

Barciszewska, M., G. Keith, G. Dirheimer and J. Barciszewski, 1988 The nucleotide-sequence of transfer $RNA^{ser}$ and transfer $RNA^{ile}$ of yellow lupin seeds. Nucleic Acids Res. **16:** 8175.

Bartolome, C., X. Maside and B. Charlesworth, 2002 On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster.* Mol. Biol. Evol. **19:** 926–937.

Beguiristain, T., M. A. Grandbastien, P. Puigdomenech and J. M. Casacuberta, 2001 Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol. **127:** 212–221.

Bennett, M. D., A. V. Cox and I. J. Leitch, 1998 Angiosperm DNA C-values database. http://www.rbgkew.org.uk/cval/database1.html.

Bennetzen, J. L., 2002 Mechanisms and rates of genome expansion and contraction in flowering plants. Genetica **115:** 29–36.

Bowen, N. J., and J. F. McDonald, 1999 Genomic analysis of *Caenorhabditis elegans* reveals ancient families of retroviral-like elements. Genome Res. **9:** 924–935.

Burke, J. M., T. J. Voss and M. L. Arnold, 1998 Genetic interactions and natural selection in Louisiana iris hybrids. Evolution **52:** 1304–1310.

Carr, M., J. R. Soloway, T. E. Robinson and J. F. Y. Brookfield, 2002 Mechanisms regulating the copy numbers of six LTR retrotransposons in the genome of *Drosophila melanogaster.* Chromosoma **110:** 511–518.

Casacuberta, J. M., S. Vernhettes and M. A. Grandbastien, 1995 Sequence variability within the tobacco retrotransposon Tnt1 population. EMBO J. **14:** 2670–2678.

Casacuberta, J. M., S. Vernhettes, C. Audeon and M. A. Grandbastien, 1997 Quasispecies in retrotransposons: a role for sequence variability in Tnt1 evolution. Genetica **100:** 109–117.

Charlesworth, B., and C. H. Langley, 1989 The population genetics of Drosophila transposable elements. Annu. Rev. Genet. **23:** 251–287.

Charlesworth, D., and S. I. Wright, 2001 Breeding systems and genome evolution. Curr. Opin. Genet. Dev. **11:** 685–690.

Curcio, M. J., and D. J. Garfinkel, 1999 New lines of host defense: inhibition of Ty1 retrotransposition by Fus3p and NER/TFIIH. Trends Genet. **15:** 42–45.

Doyle, J. J., and J. L. Doyle, 1987 A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. **19:** 11–15.

Feschotte, C., J. Jiang and S. R. Wessler, 2002 Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. **3:** 329–341.

Flavell, A., E. Dunbar, R. Anderson, S. Pearce, R. Hartley *et al.*, 1992 *Ty1-copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res. **20:** 3639–3644.

Flavell, R. B., M. D. Bennett, J. B. Smith and D. B. Smith, 1974 Genome size and proportion of repeated nucleotide-sequence DNA in plants. Biochem. Genet. **12:** 257–269.

Fu, H. H., and H. K. Dooner, 2002 Intraspecific violation of genetic colinearity and its implications in maize. Proc. Natl. Acad. Sci. USA **99:** 9573–9578.

Galbraith, D. W., G. M. Lambert, J. Macas and J. Dolezel, 1997 Analysis of nuclear DNA content and ploidy in higher plants, pp. 7.6.1–7.6.22 in *Current Protocols in Cytometry,* edited by J. P. Robinson, Z. Darzynkiewicz, P. Dean, A. Orfao, P. Rabinovitch, C. C. Stewart, H. Tanke and L. Wheeless. John Wiley & Sons, New York.

Grandbastien, M.-A., 1998 Activation of plant retrotransposons under stress conditions. Trends Plant Sci. **3**: 181–187.

Hamrick, J. L., and M. J. W. Godt, 1996 Effects of life history traits on genetic diversity in plant species. Philos. Trans. R. Soc. Lond. B Biol. Sci. **351:** 1291–1298.

Hull, R., 1999 Classification of reverse transcribing elements: a discussion document. Arch. Virol. **144:** 209–214.

Johnston, J. S., M. D. Bennett, A. L. Rayburn, D. W. Galbraith and H. J. Price, 1999 Reference standards for determination of DNA content of plant nuclei. Am. J. Bot. **86:** 609–613.

Joseph, J. L., J. W. Sentry and D. R. Smyth, 1990 Interspecies distribution of abundant DNA sequences in Lilium. J. Mol. Evol. **30:** 146–154.

Kalendar, R., J. Tanskanen, S. Immonen, E. Nevo and A. H. Schulman, 2000 Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE*-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proc. Natl. Acad. Sci. USA **97:** 6603–6607.

Kidwell, M. G., 2002 Transposable elements and the evolution of genome size in eukaryotes. Genetica **115:** 49–63.

Kimura, Y., Y. Tosa, S. Shimada, R. Sogo, M. Kusaba *et al.*, 2001 OARE-1, a Ty1-copia retrotransposon in oat activated by abiotic and biotic stresses. Plant Cell Physiol. **42:** 1345–1354.

Kumar, A., and J. L. Bennetzen, 1999 Plant retrotransposons. Annu. Rev. Genet. **33:** 479–532.

Labrador, M., M. Farre, F. Utzet and A. Fontdevila, 1999 Interspecific hybridization increases transposition rates of Osvaldo. Mol. Biol. Evol. **16:** 931–937.

Leeton, P. R. J., and D. R. Smyth, 1993 An abundant LINE-like element amplified in the genome of *Lilium speciosum*. Mol. Gen. Genet. **237:** 97–104.

Levin, H. L., 2002 Newly identified retrotransposons of the Ty3/*gypsy* class in fungi, plants, and vertebrates, pp. 684–701 in *Mobile DNA II*, edited by N. L. Craig, R. Craigie, M. Gellert and A. M. Lambowitz. American Society for Microbiology Press, Washington, DC.

Malik, H. S., and T. H. Eickbush, 1999 Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. J. Virol. **73:** 5186–5190.

Malik, H. S., S. Henikoff and T. H. Eickbush, 2000 Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses. Genome Res. **10:** 1307–1318.

Manninen, I., and A. H. Schulman, 1993 Bare-1, a Copia-like retroelement in barley (*Hordeum vulgare* L). Plant Mol. Biol. **22:** 829–846.

Marin, I., and C. Llorens, 2000 Ty3/Gypsy retrotransposons: description of new *Arabidopsis thaliana* elements and evolutionary perspectives derived from comparative genomic data. Mol. Biol. Evol. **17:** 1040–1049.

McClintock, B., 1984 The significance of responses of the genome to challenge. Science **226:** 792–801.

Meyers, B. C., S. V. Tingley and M. Morgante, 2001 Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. Genome Res. **11:** 1660–1676.

Nakajima, R., K. Noma, H. Ohtsubo and E. Ohtsubo, 1996 Identification and characterization of two tandem repeat sequences (TrsB and TrsC) and a retrotransposon (RIRE1) as genome-general sequences in rice. Genes Genet. Syst. **71:** 373–382.

Preston, B. D., 1996 Error-prone retrotransposition: rime of the ancient mutators. Proc. Natl. Acad. Sci. USA **93:** 7427–7431.

Pringle, C. R., 1999 Virus taxonomy—1999—the universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. Arch. Virol. **144:** 421–429.

Randolph, L. F., 1966 *Iris nelsonii*, a new species of Louisiana iris of hybrid origin. Baileya **14:** 143–169.

Randolph, L. F., J. Mitra and I. S. Nelson, 1961 Cytotaxonomic studies of Louisiana irises. Bot. Gaz. **123:** 126–131.

Randolph, L. F., I. S. Nelson and R. L. Plaisted, 1967 Negative evidence of introgression affecting the stability of Louisiana iris species. Mem. Cornell Univ. Agric. Exp. Stn. **398:** 1–56.

Riley, H. P., 1938 A character analysis of colonies of *Iris fulva*, *Iris hexagona* var. *giganticaerulea* and natural hybrids. Am. J. Bot. **24:** 649–725.

Robinson, T. J., O. Wittekindt, J. J. Pasantes, W. S. Modi, W. Schempp *et al.*, 2000 Stable methylation patterns in interspecific antelope hybrids and the characterization and localization of a satellite fraction in the Alcelaphini and Hippotragini. Chromosome Res. **8:** 635–643.

Roemer, I., F. Grutzner, H. Winkig, T. Haff, A. Orth *et al.*, 1999 Global methylation in eutherian hybrids. Nature **40:** 131–132.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

SanMiguel, P., and J. 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.

SanMiguel, P., A. Tikhonov, J. Young-Kwan, N. Motchoulakaia, P. S. Springer *et al.*, 1996 Nested retrotransposons in the intergenic regions of the maize genome. Science **274:** 765–768.

SanMiguel, P., B. S. Gaut, A. Tikhonov, Y. Nakajima and J. L. Bennetzen, 1998 The paleontology of intergene retrotransposons of maize. Nat. Genet. **20:** 43–45.

Seeley, K. A., D. H. Byrne and J. T. Colbert, 1992 Red light-independent instability of oat phytochrome messenger-RNA *in vivo*. Plant Cell **4:** 29–38.

Soltis, D. E., P. S. Soltis, T. G. Collier and M. L. Edgerton, 1991 Chloroplast DNA variation within and among genera of the Heuchera group (Saxifragaceae): evidence for chloroplast transfer and paraphyly. Am. J. Bot. **78:** 1091–1112.

Suoniemi, A., A. Narvanto and A. Schulman, 1996 The *BARE*-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol. Biol. **31:** 295–306.

Van den Broeck, D., T. Maes, M. Saur, J. Zetof, P. De Keukeleire *et al.*, 1998 *Transposon Display* identifies individual transposable elements in high copy number lines. Plant J. **13:** 121–129.

Vernhettes, S., M.-A. Grandbastien and J. M. Casacuberta, 1998 The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. Mol. Biol. Evol. **15:** 827–836.

Vicient, C. M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav *et al.*, 1999 Retrotransposon *BARE-1* and its role in genome evolution in the genus *Hordeum*. Plant Cell **11:** 1769–1784.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee *et al.*, 1995 AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. **23:** 4407–4414.

Waugh, R., K. McLean, A. J. Flavell, S. R. Pearce, A. Kumar *et al.*, 1997 Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. **253:** 687–694.

Waugh O'Neil, R. J., M. J. O'Neil and J. A. Marshall Graves, 1998 Undermethylation associated with retroelement activation and chromosome remodeling in a interspecific mammalian hybrid. Nature **393:** 68–72.

Wright, D. A., and D. F. Voytas, 1998 Potential retroviruses in plants: Tat1 is related to a group of *Arabidopsis thaliana* Ty3/gypsy retrotransposons that encode envelope-like proteins. Genetics **149:** 703–715.

Communicating editor: J. A. Birchler