Copyright © 2003 by the Genetics Society of America

An *Ac*-like Transposable Element Family With Transcriptionally Active Y-Linked Copies in the White Campion, *Silene latifolia*

Ellen J. Pritham,*[,1,2] Y. Hi Zhang,*[,1] Cédric Feschotte† and Rick V. Kesseli*[,3]

*Department of Biology, University of Massachusetts, Boston, Massachusetts 02125 and †Department of Plant Biology, University of Georgia, Athens, Georgia 30602

Manuscript received April 3, 2003
Accepted for publication June 5, 2003

ABSTRACT

An RFLP genomic subtraction was used to isolate male-specific sequences in the species *Silene latifolia*. One isolated fragment, SLP2, shares similarity to a portion of the *Activator* (*Ac*) transposase from *Zea mays* and to related proteins from other plant species. Southern blot analysis of male and female *S. latifolia* genomic DNA shows that SLP2 belongs to a low-copy-number repeat family with two Y-linked copies. Screening of a *S. latifolia* male genomic library using SLP2 as a probe led to the isolation of five clones, which were partially sequenced. One clone contains two large open reading frames that can be joined into a sequence encoding a putative protein of 682 amino acids by removing a short intron. Database searches and phylogenetic analysis show that this protein belongs to the hAT superfamily of transposases, closest to Tag2 (*Arabidopsis thaliana*), and contains all of the defined domains critical for the activity of these transposases. PCR with genomic and cDNA templates from *S. latifolia* male, female, and hermaphrodite individuals revealed that one of the Y-linked copies is transcriptionally active and alternatively spliced. This is the first report of a transcriptionally active transposable element (TE) family in *S. latifolia* and the first DNA transposon residing on a plant Y chromosome. The potential activity and regulation of this TE family and its use for Y chromosome gene discovery is discussed.

THE white campion, *Silene latifolia* (previously *Melandrium album*), is a common eudicot weed found in both North America and Europe. There are ~700 species in the genus Silene, the majority of which are outcrossing and hermaphroditic. Breeding systems range, however, from nearly obligate selfing (cleistogamy) to obligate outcrossing (dioecy). Dioecy (separate male and female individuals) likely evolved at least twice from a hermaphroditic state within the last 20–30 million years (Desfeux *et al.* 1996).

S. latifolia is dioecious with an XY-chromosome-based sex determination system analogous to the mammalian system. Dioecy is stable, although hermaphroditic mutants (XY) occur at extremely low frequency in natural populations (our personal observations) and can be induced through the exogenous application of demethylating agents such as 5-azacytidine and with gamma irradiation (Donnison *et al.* 1996; Janousek *et al.* 1996). Dioecy is a relatively rare breeding system in plants and is not necessarily coupled with sexually dimorphic chromosomes. The recent evolution of dioecy in *S. latifolia* from a hermaphroditic ancestor distinguishes this system from sex determination in animal systems where chromosome-based dioecy has existed for 200–300 million years (Lahn and Page 1999). This provides the unique opportunity to isolate the genetic loci necessary for the development of separate sexes and to trace some of the chromosomal changes that have occurred through comparisons with hermaphroditic species. Therefore, *S. latifolia* is a model for studying the processes underlying the separation of gender and the evolution of dimorphic sex chromosomes in plants (for review see Charlesworth and Guttman 1999; Moneger 2001; Charlesworth 2002).

The Y chromosome of *S. latifolia* is the largest chromosome (~925 Mb) in the genome (Westergaard 1946) and is ~50% larger than the X chromosome, in contrast to the mammalian system, where the Y chromosome is typically dramatically smaller. This difference in chromosome size has not yet been explained, although transposable element movement and accumulation is one potential cause. The lack of recombination over the majority of the Y chromosome is thought to reduce the efficacy of selection (for review see Charlesworth 2002) and thus allow transposable elements with potentially detrimental fitness impacts to accumulate in these nonrecombining chromosomal regions (Steinemann and Steinemann 1991, 1992, 1993, 1997, 1998, 2000; Steinemann *et al.* 1993; Hochstenbach *et al.* 1994; Abe *et al.* 1998; Esposito *et al.* 1999; Underwood and Bianco 1999; Erlandsson *et al.* 2000). Retrotransposon-like se-

Sequence data from this article have been deposited in the EMBL/GenBank Data Libraries under accession nos. AY172035–AY172052.

[1]These authors contributed equally to this article.

[2]*Present address:* Department of Plant Biology, University of Georgia, Athens, GA 30602.

[3]*Corresponding author:* University of Massachusetts, Biology Department, 100 Morrissey Blvd., Boston, MA 02125.
E-mail: rick.kesseli@umb.edu

quences have been isolated previously during screens for Y-linked sequences in *S. latifolia* (Donnison *et al.* 1996; Nakao *et al.* 2002) and in other dioecious plant species (Sakamoto *et al.* 2000; Shibata *et al.* 2000).

Here we have used restriction fragment length polymorphism (RFLP) genomic subtraction to isolate male-specific sequences from the progeny of an intraspecific cross of *S. latifolia*. This technique was used to isolate low-copy sequences linked to the Y chromosome. One of the fragments, SLP2, was similar to the maize *Activator* (*Ac*) transposase and belonged to a low-copy-number repeat family with two Y-linked copies. We found that an intact transposase gene still resides in the genome of *S. latifolia*, but probably not on the Y chromosome. We provide evidence, however, that one of the Y-linked transposase pseudogenes is transcribed and alternatively spliced in both male and hermaphrodite leaf and bud tissues. These results are discussed in the context of Y chromosome evolution in *S. latifolia*, with an emphasis on the potential activity of this transposable element family and its use as a tool for Y-linked gene discovery.

MATERIALS AND METHODS

Plant material: The F_1 population was generated using an individual from an Italian population as the female parent and an individual from a western Massachusetts population as the male parent. The hermaphroditic plant (chromosomally XY) was derived from a single hermaphroditic mutant found in a wild population at the University of Massachusetts, Boston, and maintained by crossing wild-type female plants with hermaphroditic pollen or selfing the hermaphroditic individual. Because Y chromosome deletions can yield hermaphrodites (Westergaard 1946), this individual was used here in comparative screens with normal males. All plant materials were maintained in the greenhouse at the University of Massachusetts, Boston.

RFLP subtraction and Southern hybridization: This procedure is a modification of the RFLP subtraction described in Rosenberg *et al.* (1994). Female DNA sequences were subtracted from male DNAs, allowing the isolation of male-specific sequences located on the nonrecombining portion of the Y chromosome. The "tester" DNA was extracted from an F_1 male individual (Italy female × western Massachusetts male) and the "driver" DNA was extracted from a pool of 10 female individuals from the same cross. Oligonucleotides were purchased from Amitof Biotech (Boston). The male-specific bands were gel purified with a QIA gel extraction kit (QIAGEN, Chatsworth, CA). The product was ligated to *Hin*dIII-digested dephosphorylated pUC 18 vector. The insert DNAs from 24 white clones were prepared individually by PCR using pUC18/M13 primers and screened. This DNA was used as the template to make radioactive probes ^{32}P-labeled for hybridizing to *Hin*dIII-digested pooled male and female genomic DNA. Positive clones were labeled and hybridized to the pooled male and female genomic DNA for verification.

Genomic DNAs for Southern analysis were digested with *Hin*dIII at 3 units/μg DNA for 3 hr at 37°. The digested DNAs were separated on a 1% agarose gel. Southern transfer and hybridization followed (Bernatzky and Tanksley 1986).

λ-Library construction and screening: The genomic DNA was extracted from a pool of five male F_1 individuals from the previously described cross and partially digested with *Sau*3A. The digested DNAs were separated on a 1% agarose gel and the 9- to 23-kb DNA fragment was purified with QIAEXII gel extraction kit. The purified DNA was ligated to λ-arms from the Stratagene (La Jolla, CA) Lambda DASH II/*Bam*HI kit. The cloning and screening were performed following the company's protocol. The library was screened with SLP2 (previously described) from this study. Positive clones were picked and DNA was extracted.

To determine the insert size of positive clones, field inversion gel electrophoresis was performed using the PC500 switchback pulse controller system (Pharmacia, Piscataway, NJ). Gel electrophoresis was performed in 1% Seakem LE agarose gel in 0.5× TBE with the following pulse conditions: 150 V, $F:R$ = 3:1, pulse time = 0.6–2.0 sec for 24 hr.

Primer design and PCR walking: The primers used in PCR walking were designed using a computer program Oligo 3.0 and synthesized by Operon Technologies (Alameda, CA). PCR walking followed the principle previously described (Parker *et al.* 1991). For each PCR, a specific primer (18–24 bases) and a random primer (18–24 bases) were used. DNA prepared from positive clones screened by SLP2 served as the template in PCR walking. Specific primers were designed using the sequence from SLP2 in an attempt to retrieve sequences allowing the reconstruction of the coding region. The PCR mixture was 2.5 μl of 10× buffer (100 mm Tris-HCl, pH 8.3, 500 mm KCl, 15 mm $MgCl_2$, 0.1% gelatin), 2.0 μl of dNTP (1.25 mm each of dATP, dCTP, dGTP, and dTTP), 0.5 μl of specific and random primer (10 mm), 2.0 μl of $MgCl_2$ (10 mm), 0.2 μl of *Taq* DNA polymerase (5 units/μl), 2.5 μl of DNA (5 ng/μl), and 14.8 μl of H_2O. The cycling parameters were 1 cycle 94° 1 min; 30 cycles 94° 1 min; 50° 1 min; 72° 2 min; 1 cycle 72° 5 min, 4° hold.

Sequencing, database searches, and mining: Clones and PCR products were sequenced using d-rhodamine dye-terminator or Big Dye chemistry with an ABI prism 377 from Applied Biosystems (Foster City, CA). Homology searches were completed using BLASTX and the nonredundant databases at the National Center for Biotechnology Information (NCBI) website (Gish and States 1993). Intron prediction was completed using the NetGene2 server (Hebsgaard *et al.* 1996). Closely related transposase sequences were gathered from the database through reiterative searches with the partial sequence of one λ-clone and closely related accessions using the search algorithm BLASTX (Gish and States 1993).

Sequence comparison and phylogeny construction: DNA or protein sequences were aligned using the software package ClustalW (Thompson *et al.* 1994). Pairwise comparisons were made with the Blast2 Sequence function at NCBI (Altschul *et al.* 1997). The phylogenies were constructed using parsimony or distance methods with 1000 bootstrap replicates of the consensus and the programs provided in the PAUP* package, Version 4 (Swofford 2002). Branches from parsimony or distance trees were labeled with bootstrap values on the well-supported branches of interest. Well-supported branches from the *S. latifolia*, which are composed of genomic-derived sequences from multiple individuals, are considered to represent unique genomic locations.

RT-PCR: RNA was extracted using tissue collected from plants provided by Paige Dennis (University of Massachusetts, Boston). Tissue was ground in liquid nitrogen and the RNeasy Mini RNA extraction kit available from QIAGEN. Total RNA was treated with DNaseI purchased from GIBCO BRL (Gaithersburg, MD), 1 unit DNase I per 1 μg RNA, incubated at room temperature for 15 min and EDTA inactivated. Reverse transcription was completed using 3′ rapid amplification of cDNA ends. SUPERSCRIPT II RNase H-reverse transcriptase was purchased from GIBCO BRL. An RT− control was com-

pleted for each sample. The primer sequence for the reverse transcription reaction is: RT Anchor, 5′-CCGACACGACGAC TACAGCATTTTTTTTTTTTTTTTTTTTTV-3′. RNA was incubated with 50 μM of RT Anchor primer at 70° for 10 min. The reaction was completed according to the manufacturer's protocol. Two rounds of PCR were completed using a nested primer strategy. The following primers were used: RT Amp, 5′-CCGACACGACGACTACAGCA- 3′; intron F, 5′-CATGGCT GACAGAATGAGGATC-3′; intron R, 5′-AAGAGTGTTGTCC TCCATTTCATC-3′. The first round of PCR was completed using the primers RT Amp and intron F. The second round of PCR utilized the primers intron F and intron R. The primers were used at a final concentration of 1 μM. Amplitaq Gold brand *Taq* from Applied Biosystems was used at a final concentration of 0.25 units at the suggestion of the manufacturer. The cycling parameters were step 1, 10 min at 95°; step 2, 1 min 95°; step 3, 1 min 52°; step 4, 1 min 72°; step 5, 9 more cycles to 2; step 6, 1 min 95°; step 7, 1 min 55°; step 8, 1 min 72°; step 9, 25 more cycles to 5; step 10, 5-min extension at 72°. The PCR product was separated on a 1.2% agarose gel 1× TBE and the two resulting bands were excised and gel purified using the kit previously described.

Genomic PCR and cloning: Genomic DNA from one female, one male, and one hermaphroditic (chromosomally XY) individual was used as the template for PCR. These individuals were the same as those used for the RT-PCR. The intron F and intron R primers and PCR conditions were used as described above. PCR products were purified directly using a QIAquick PCR purification kit (QIAGEN). PCR products were ligated into the PGEM-T vector (Promega, Madison, WI) as directed by the manufacturer's protocol. XL1 blue cells (Stratagene) were transformed with plasmids and screened with blue/white screening. Between 5 and 25 white colonies were picked from each cloning procedure. Plasmids were prepared for sequencing using a Qia-prep spin miniprep kit (QIAGEN) or Templiphi (Amersham, Buckinghamshire, UK).

RESULTS

Isolation and identification of male-specific sequences: Fifty-eight recombinant clones obtained from one transformation of subtraction products were transferred to a nylon membrane and probed with pooled female genomic DNA. On the basis of the intensity of the hybridization signals, 24 clones appeared to contain only low- or single-copy DNA fragments (data not shown).

From a F_1 population, genomic DNAs from five males and five females were digested with *Hin*dIII and probed with the plasmid inserts isolated in the RFLP subtraction. Fragments present in males but absent from females are likely to be linked to the nonrecombining portion of the Y chromosome. One clone, SLP2, hybridized to 4–8 bands in both males and females, including two unique male-specific fragments (Figure 1). The male-specific fragments are ∼3.0 and 0.9 kb in size. In addition, a band of ∼0.7 kb is twice as bright in all female individuals when compared to all males, indicating that these fragments are likely X-linked, although other explanations are possible. Additional fragments of ∼16 and 7 kb in the SLP2 hybridization show 1:1 segregation patterns in males and females, indicating that they are located to either maternal X chromosomes or autosomes.

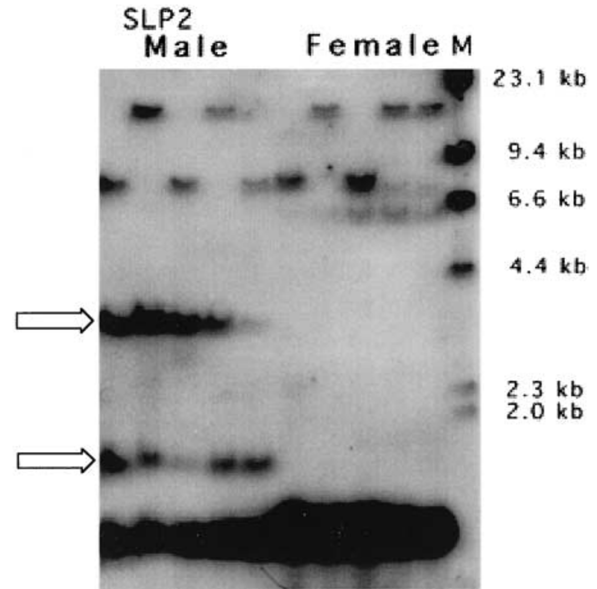

FIGURE 1.—Southern blot of F_1 male and female genomic DNA digested with *Hin*dIII and hybridized with SLP2. The arrows indicate the two male-specific fragments. The size marker (M) is λ cut with *Hin*dIII.

SLP2 has similarity to the *Ac* transposase: The 827-bp SLP2 insert was sequenced and used as a query to search the nonredundant database at NCBI using BLASTX with default settings. This search revealed that the entire sequence of SLP2 shares significant similarity to the transposases from the hAT (*hobo/Ac/Tam3*) superfamily (*E* values for the first 45 matches ranged from 1×10^{-64} to 8×10^{-04}), including those encoded by the *Ac* element from *Zea mays* (E value $= 5 \times 10^{-41}$). This fragment encodes the first three of the six conserved hAT superfamily domains (Rubin *et al.* 2001).

Isolation of SLP2-related sequences from a male *S. latifolia* genomic library and sequence analysis: A genomic λ-library was constructed from a pool of five male full siblings. Screening of the library (one million plaques) using SLP2 as a probe at high stringency led to the isolation of five unique genomic clones. On the basis of pulsed-field gel electrophoresis, the size of the inserts of these five clones ranged from ∼12 to 22 kb. Primers were designed on the basis of the sequence of SLP2 to allow the amplification of this region from the λ-clones. Restriction digests of the clones and sequence comparisons of the SLP2 region showed that each clone was unique. These clones, designated Thelma2, Thelma6-1, Thelma6-2, Thelma7, and Thelma13 (Figure 2), were 93–100% identical to the 827-bp region of SLP2. Thelma7 is the only sequence identical to SLP2 and thus is a genomic clone with an insert that likely originated from the Y chromosome.

Primer walking was used to sequence the regions flanking SLP2 for two of the λ-clones, Thelma7 and Thelma13. A consensus was based on a sequence align-

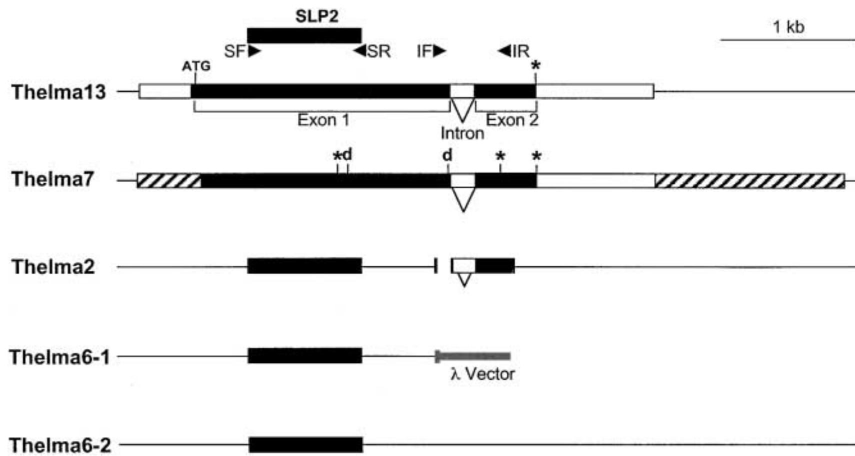

FIGURE 2.—The regions sequenced (boxed) of the λ-clones flanking the original male-specific plasmid SLP2. SLP2 was used as a probe to a male λ genomic library and five clones, Thelma13, -7, -2, -6-2, and -6-1 (12–22 kb), were recovered. Primers (SF and SR are indicated by arrowheads) based on the sequence of SLP2 were used to amplify the corresponding region from each of the λ-clones. The 3482-bp Thelma13 fragment containing the complete transposase coding region (coding region indicated by solid box) was sequenced by primer walking using primers designed to SLP2 and random primers. The region flanking the putative intron was amplified using primers (IF and IR indicated by arrowheads) from Thelma7, -13, and -2. The 5174-bp region of Thelma7 was sequenced (regions that are not homologous to Thelma13 are indicated by diagonal lines). Multiple stop codons (*) and deletions (d) resulting in frameshifts indicate that it is a likely pseudogene.

ment of at least five overlapping fragments and resulted in a 3483-bp sequence for Thelma13 and a 5174-bp sequence for Thelma7. This includes the region corresponding to SLP2 and extends in both directions (Figure 2). An alignment of the Thelma13 and Thelma7 fragments revealed that they are 94% identical over a 2948-bp region with 11 gaps. The alignment begins in the 5′ region of Thelma13 at position 534 bp and extends to the end. This region aligns at position 1036 of Thelma7. The two genomic regions were analyzed in detail using a translation of all six reading frames and intron prediction tools. This analysis revealed that Thelma13 contains two large open reading frames (ORFs) of 549 and 133 amino acids (aa) that can be joined together by the removal of a 135-bp intron, predicted with 94% confidence by NetGene (HEBSGAARD *et al.* 1996). The resulting protein is 682 aa in length, a size similar to various transposases from the hAT superfamily. Analysis of Thelma7 using the ORFs defined in Thelma13 indicates that Thelma7 is a pseudogene, with a likely insertion or deletion in the 5′ end of the sequence. No obvious structural characteristics of class 2 transposable elements were detected in the sequences flanking the coding regions.

Transposase domain comparisons: Predicted proteins were mined from the nonredundant database through reiterative homology searches with the deduced protein sequence from Thelma13 and related accessions. Retrieved protein sequences were aligned and a phylogenetic tree was constructed using the neighbor-joining method. This analysis reveals that the Thelma13 putative transposase belongs to the *Ac/Tam3* clade of the hAT superfamily of transposases. It is most closely related to the transposase encoded by the Tag2 element and other undescribed putative transposases from *Arabidopsis thaliana* (Figure 3).

The putative transposases from Thelma13, Tag2, and *Ac* (ORFa) were aligned to illustrate the close relationship between these proteins and therefore the likely conservation of function (Figure 4). The six hAT superfamily domains defined by (RUBIN *et al.* 2001) are present in each of these putative transposase proteins; the biochemical function of all of these domains is not yet known. The N terminus of Thelma13 also shares similarity to the DNA-binding domain defined biochemically for *Ac*; this region is absent in Tag2 (see Figure 4). A BED finger DNA-binding domain (ARAVIND 2000), predicted by the conserved domain database at NCBI, is identified in *Ac* and Thelma13, but is absent in Tag2 (underlined region 2 in Figure 4).

Overall, the putative Thelma13 protein product is 29% (196/655 aa) identical and 47% similar (315/655 aa) to the *Ac* transposase. The most highly conserved domain is located in the C terminus of the protein and is involved in dimerization and likely in catalysis (ESSERS *et al.* 2000). The similar residues are present in all regions of the protein but concentrated in the domains known to be critical for function or well conserved across the superfamily. This observation strongly suggests that the Thelma13 and *Ac* transposases function in a similar manner.

Transcriptional analysis: With primers flanking the single predicted intron, we used RT/PCR to amplify cDNA from single male, female, and hermaphroditic individuals. Two bands of ~500 and 600 bp, respectively, were visualized by agarose electrophoresis in both the XY male and XY hermaphrodite, but not in the XX female (Figure 5a). These bands were excised, cloned, and multiple clones were sequenced. The sequence of the smaller fragment revealed that the intron had been correctly predicted and removed. The sequence of the larger fragment was identical except that it retained the intron. Thus, the confirmed intron appears to be alternatively spliced, resulting in the presence of two transcripts originating from one element. Both transcripts are found simultaneously in leaf and bud tissue

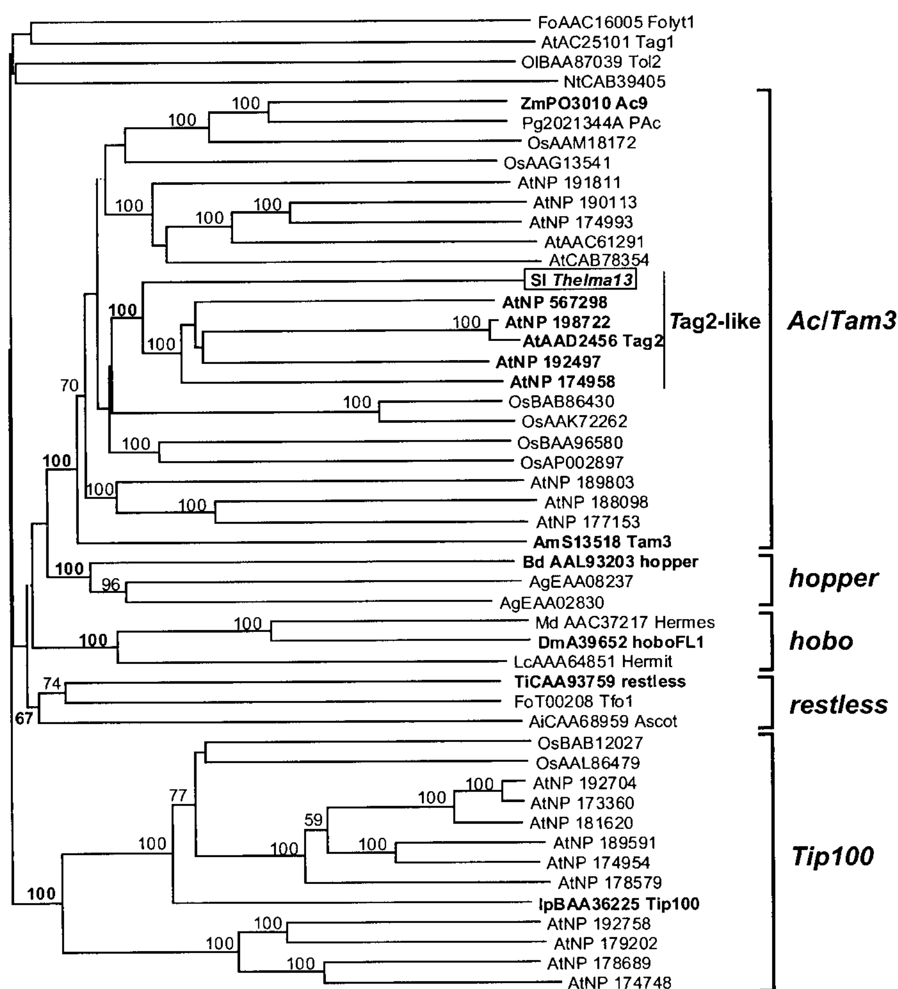

FIGURE 3.—Comparative phylogenetic analysis of hAT superfamily proteins with the putative Thelma13 transposase. The sequences are labeled with their GenBank accession numbers. The first two letters are the genus species abbreviations (Fo, *Fusarium oxysporum*; At, *A. thaliana*; Ol, *Oryzias latipes*; Nt, *Nicotiana tabacum*; Zm, *Z. mays*; Pg, *Pennisetum glaucum*; Os, *Oryza sativa*; Sl, *S. latifolia*; Am, *Antirrhinum majus*; Bd, *Bactrocera dorsalis*; Ag, *Anopheles gambiae*; Md, *Musca domestica*; Dm, *Drosophila melanogaster*; Lc, *Lucilia cuprina*; Ti, *Tolypocladium inflatum*; Ai, *Ascobolus immersus*; Ip, *Ipomoea purpurea*). The transposable elements that have been described are named after the accession number. The well-supported clades are indicated by brackets and given the name of the first described element in that clade.

from the male and hermaphrodite (both XY), but not from the female. The appropriate controls eliminated the likelihood of genomic contamination.

The primers flanking the intron site were also used to amplify genomic DNA from the male, female, and hermaphroditic individuals. A single sharp band of ~600 bp was obtained from each individual (data not shown). This band was directly cloned and 20–25 clones were sequenced for each individual, with the goal of identifying most of the amplifiable copies in each of these genomes and the specific copy being expressed. Six of 25 copies from the male, 3 of 20 copies from the female, and 2 of 20 copies from the hermaphrodite were unique. We constructed a phylogeny of the unique amplified sequences (per individual), excluding the intron (Figure 5b), to understand the relationship of these sequences and to estimate copy number in the genome. The phylogeny includes the cDNA fragments from male and hermaphrodite; the unique genomic clones from the male, female, and hermaphrodite; and the corresponding regions of the genomic clones Thelma7, -13, and 2. The deduced phylogeny reveals three major clades supported by high bootstrap values. The phylogeny is further broken down into five subdivisions labeled A–E, each likely representing a separate locus with males having copies in all five branches. Branches A and E represent the two putative Y-linked copies first identified by Southern blots (Figure 1). Branch A consists solely of the Y-linked λ-clone Thelma7 fragment. Branch E contains one genomic clone from the male (gDNAm13) and one from the hermaphrodite (gDNAh16). The clade also contains the two alternatively spliced transcripts, RNA1m21 and RNA2m9, isolated from the male and identical to the male genomic clone, and the two alternatively spliced transcripts isolated from the hermaphrodite and identical to its genomic clone. The genomic sequence gDNAm13 is 99.2% identical to the genomic copy gDNAh16, suggesting that they are allelic. Branch E is a strongly supported clade that lacks female sequences, but includes copies with gender-specific expression. Together, this suggests that branch E is a male-specific Y-linked grouping.

DISCUSSION

We report a novel gene family in the genome of *S. latifolia*, with homology to a large superfamily of class 2 transposase proteins. Class 2 transposons are mobile

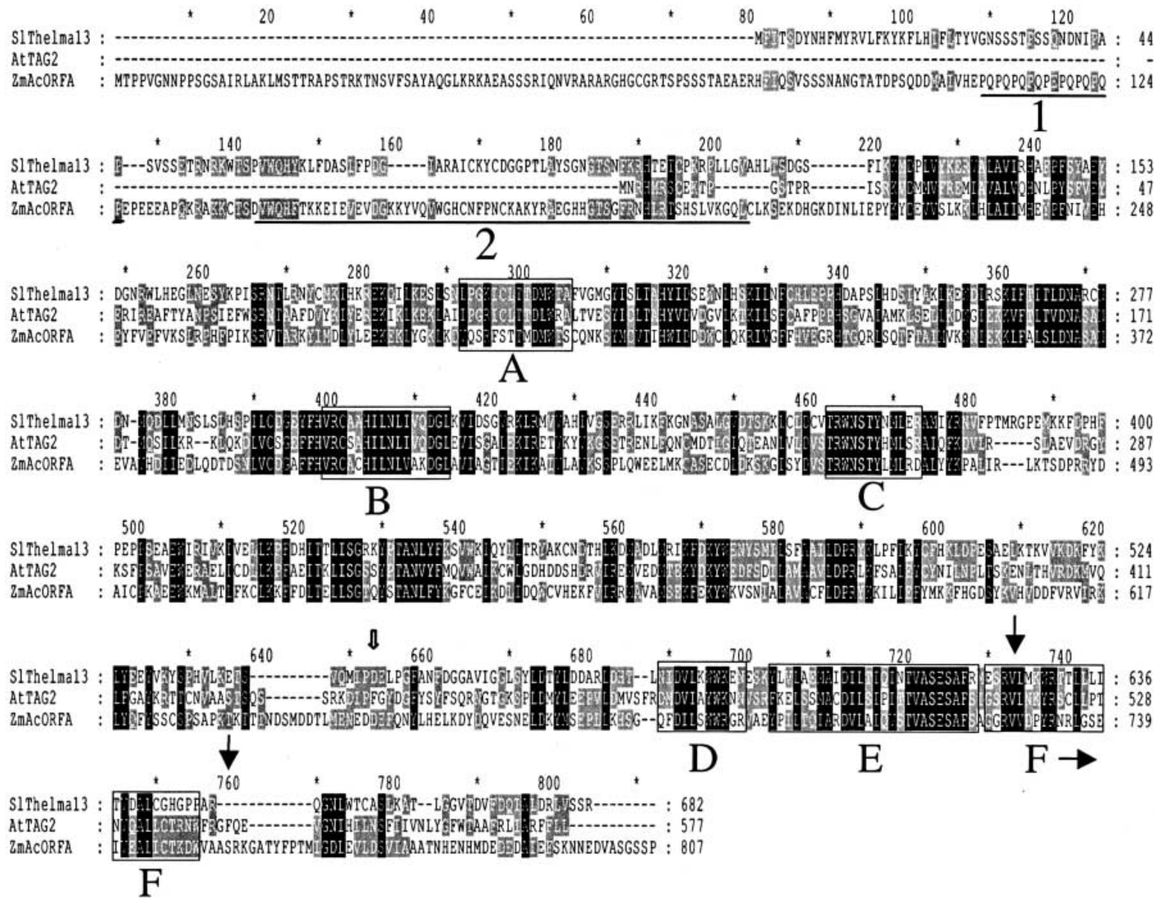

Figure 4.—Boxshade Thelma13, Tag2, and *Ac* (ORFa). The solid boxes indicate identical or similar residues shared between all sequences. The shaded boxes indicate identical or similar residues shared by two of the sequences. The first underlined region is involved in DNA localization. The second underlined region represents the BED zinc finger DNA-binding domain (Aravind 2000) conserved between *Ac* and Thelma13, predicted by RPS-BLAST at NCBI, and includes the DNA-binding domain biochemically determined for *Ac*. Conserved hAT superfamily protein blocks are indicated by boxes and represented by letters A–F (Rubin *et al.* 2001). The regions marked D–F are located in the region involved in dimerization (Essers *et al.* 2000). An open arrow identifies the position of the single alternatively spliced intron shared by Thelma and Tag2. *Ac* transposase mutants truncated at the regions marked by the solid arrows upregulate transposition when cotransfected with wild-type TPase (Kunze *et al.* 1993).

genetic units capable of changing chromosomal location, via a DNA-mediated, cut-and-paste mechanism (for review see Feschotte *et al.* 2002). An intact putative transposase coding region in *S. latifolia*, Thelma13, is likely to be linked to the X or autosomes. Thelma13 belongs to the hAT superfamily of DNA transposons, branching with the *Ac/Tam3* clade, most closely related to Tag2 of Arabidopsis (Figure 3). Thelma13 possesses all six described domains found in *Ac*-like transposases. Thus, Thelma transposase genes are likely dispersed as part of a class 2 TE family, although we currently have no evidence for their mobility. Two copies in *S. latifolia* are Y-linked, but both appear to be pseudogenes, although one is transcribed in leaf and bud tissue. Thelma7, a Y-linked copy, has acquired several mutations, including a large insertion or deletion (indel) in the 5′ end, which interrupts the coding region. The degree to which Thelma7 is degraded is consistent with it being an ancient component of the Y chromosome, although no orthologous copy was identified in our survey of the XY hermaphrodite's genome. This suggests that in the hermaphrodite, the region may have been deleted and therefore includes the female suppression loci (Westergaard 1946; Lebel-Hardenack *et al.* 2002).

Transposon regulation: Class 2 transposable element movement is likely regulated at multiple levels both by the host and in an autoregulatory fashion. Alternative splicing via occasional intron retention has been reported for Tag2 (Henk *et al.* 1999) and the alignment shows that the intron detected in Thelma is in the same position as the single intron in Tag2 (Figure 4). Intron retention, the most common form of alternative splicing (Mironov *et al.* 1999), frequently functions as a post-transcriptional regulatory mechanism with the short isoform acting as a dominant-negative regulator of the longer one (Boise *et al.* 1993; Okada *et al.* 1997; Ross

et al. 1997; Srinivasula *et al.* 1999). For Tag2 and Thelma, the retention of the intron would introduce a premature stop codon and lead to a C-terminal truncation. Thus, the translation of both transcripts would result in the presence of long and short protein isoforms. The long isoform would contain all the domains necessary for catalyzing transposition and presumably functions as the transposase. In contrast, the short isoform would retain the DNA localization signal, DNA-binding domain, and regions of conserved but unknown function (regions A, B, and C in Figure 4), but would not have the dimerization domain, which likely overlaps the catalytic domain (regions D, E, and F in Figure 4). The short isoform may be capable of nuclear localization and DNA interaction, but not capable of the dimerization mandatory for transposition. The short isoform might compete with the transposase (long isoform) dimers for DNA interaction and therefore prevent transposition in a concentration-dependent manner.

For some elements (*Ac* and Tag1), post-transcriptional regulation has been implicated because there is little, if any, correlation between transcript levels and excision frequencies (see Kunze and Weil 2002). Dominant-negative mutants have been found for *Ac*, but result from coexpression of proteins that lack the N terminus DNA-binding domain like those that arise from alternative transcription or translation start site usage (Kunze *et al.* 1993). Interestingly, proteins that lack regions of the C terminus when coexpressed with wild-type *Ac* transposase actually increase excision frequencies in transfected petunia protoplasts (Kunze *et al.* 1993). Further experiments are required to understand the potential role of the two isoforms in *S. latifolia*.

Y degradation and transposon tagging: The accumulation of transposable elements in chromosomal regions with low recombination has been well documented in animal species (*e.g.*, Steinemann and Steinemann 1998; Erlandsson *et al.* 2000). Class 1 TE fragments have been isolated in *S. latifolia* during screens for male-specific sequences (Donnison *et al.* 1996; Nakao *et al.* 2002) as well as in two other dioecious plant species, *Cannabis sativa* and *Rumex acetosa* (Sakamoto *et al.* 2000; Shibata *et al.* 2000). In animals, Y-linked, class 2 TEs have been reported much less frequently and this is the first report of a transposase-related sequence on a plant Y chromosome.

The lack of recombination over the majority of the Y chromosome probably causes the unavoidable decay of linked genes due to population genetic processes such as genetic hitchhiking, selective sweeps, and Muller's ratchet (for recent review see Charlesworth and Charlesworth 2000). Genes remaining active are most likely either new arrivals awaiting erosion or ancient and selectively advantageous. The Y-linked sequence, Thelma7, is clearly a pseudogene in the process of erosion with no detectable transcription. The other Y-linked Thelma locus recovered from both the male and the hermaphrodite is transcribed. These two Y-linked copies may thus be in different stages of degradation. We are currently screening closely related dioecious species to determine if Y linkage preceded speciation.

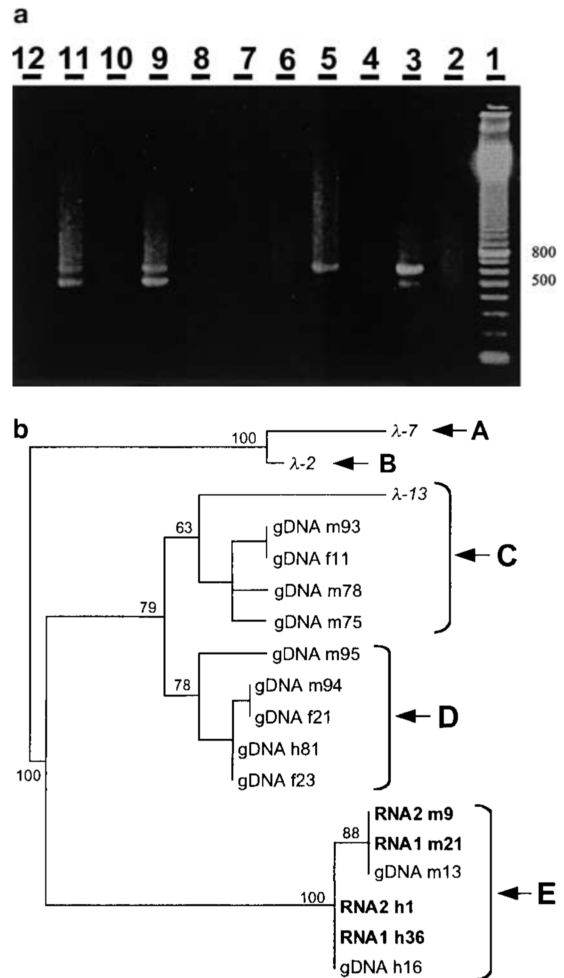

Figure 5.—(a) RT-PCR with primers designed to flank the 135-bp intron of Y-linked Thelma7. Lane 1, 100-bp-ladder size standard; lane 2, water blank; lane 3, male bud tissue; lane 4, male bud tissue without reverse transcriptase (RT−); lane 5, male leaf tissue; lane 6, male leaf tissue (RT−); lane 7, female leaf tissue; lane 8, female leaf (RT−); lane 9, hermaphrodite bud tissue; lane 10, hermaphrodite bud tissue (RT−); lane 11, hermaphrodite leaf tissue; lane 12, hermaphrodite leaf (RT−); lanes 3, 5, 9, and 11 show fragments with (≈640 bp) and without (≈500 bp) introns. (b) Phylogenetic comparison of transcribed, genomic PCR-derived, and λ-library copies corresponding to the region amplified by the primers IF and IR, but excluding the intron. Those products amplified from cDNA are labeled RNA1 or RNA2. Those products amplified from genomic DNA are labeled gDNA. RNA1 are products that were spliced. RNA2 are products that are the result of intron retention. Three major clades are well supported. The first clade includes branches A and B. A is represented by Thelma7, which is Y-linked. The second clade includes branches C and D. Included here is Thelma13, which encodes an intact transposase coding region. The third clade, branch E, possesses all the RNA-derived copies and the Y-linked genomic copies from which they are transcribed.

The Thelma family is the first class 2 transposable element family isolated in the genus Silene and the first described in plants with copies that are Y-linked. One copy not on the Y chromosome appears to be intact and capable of encoding a complete transposase with all the domains necessary to catalyze its own movement and the movement of other elements *in trans*. Previous reports of genetic XY males expressing the hermaphroditic phenotype when treated with demethylating agents (Donnison *et al.* 1996; Janousek *et al.* 1996) might be explained by the movement of hAT superfamily transposons, since methylation has been shown to be involved in their regulation (*e.g.*, *Ac* in maize reviewed in Kunze and Weil 2002). Since hAT superfamily transposable elements move both locally and into gene-rich regions (Kunze and Weil 2002), an endogenous hAT transposon tagging system may be a useful tool for gene discovery of Y-linked genes in the *S. latifolia* genome. The *S. latifolia* Y chromosome is the largest of the 24 chromosomes present in this species and is largely nonrecombining, making gene discovery by conventional methods extremely difficult. The possibility of using an endogenous transposable element family for a targeted tagging approach is worthy of further investigation, as it may be possible to bypass the need for generating transformed plants, a technically difficult procedure in Silene.

We thank Jeff Dole for supplying the individual used to establish our hermaphroditic population at the University of Massachusetts, Boston; Paige Dennis for supplying plant material and genomic DNA; Shellie Bench (Amersham BioSciences), David Peters (MJ Research), and Amy Avery for technical support; and Susan Wessler for providing facilities in support of the writing of the manuscript. This work was supported in part by the National Science Foundation (NSF), Division of Environmental Biology (93-07823), by an NSF Academic Research Infrastructure-Instrumentation grant (95-12341), and by an NSF Graduate Research Training Grant in Environmental Biology (95-52836).

LITERATURE CITED

Abe, H., M. Kanehara, T. Terada, F. Ohbayashi, T. Shimada *et al.*, 1998 Identification of novel random amplified polymorphic DNAs (RAPDs) on the W chromosome of the domesticated silkworm, *Bombyx mori*, and the wild silkworm, *B. mandarina*, and their retrotransposable element-related nucleotide sequences. Genes Genet. Syst. **73:** 243–254.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Aravind, L., 2000 The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem. Sci. **25** (9): 421–423.

Bernatzky, R., and S. Tanksley, 1986 Genetics of actin related sequences in tomato. Theor. Appl. Genet. **72:** 314–321.

Boise, L. H., M. Gonzalez-Garcia, C. E. Postema, L. Ding, T. Lindsten *et al.*, 1993 bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. Cell **74:** 597–608.

Charlesworth, B., and D. Charlesworth, 2000 The degeneration of Y chromosomes. Philos. Trans. R. Soc. Lond. B Biol. Sci. **355:** 1563–1572.

Charlesworth, D., 2002 Plant sex determination and sex chromosomes. Heredity **88:** 94–101.

Charlesworth, D., and D. S. Guttman, 1999 The evolution of dioecy and plant sex chromosome systems, pp. 25–49 in *Sex Determination in Plants*, edited by C. Ainsworth. BIOS, Oxford.

Desfeux, C., S. Maurice, J.-P. Henry, B. Lejeune and P.-H. Gouyon, 1996 Evolution of reproductive systems in the genus *Silene*. Proc. R. Soc. Lond. Ser. B **263:** 409–414.

Donnison, I. S., J. Siroky, B. Vyskot, H. Saedler and S. R. Grant, 1996 Isolation of Y chromosome-specific sequences from *Silene latifolia* and mapping of male sex-determining genes using representational difference analysis. Genetics **144:** 1893–1901.

Erlandsson, R., J. F. Wilson and S. Paabo, 2000 Sex chromosomal transposable element accumulation and male-driven substitutional evolution in humans. Mol. Biol. Evol. **17:** 804–812.

Esposito, T., F. Gianfrancesco, A. Ciccodicola, L. Montanini, S. Mumm *et al.*, 1999 A novel pseudoautosomal human gene encodes a putative protein similar to *Ac*-like transposases. Hum. Mol. Genet. **8:** 61–67.

Essers, L., R. H. Adolphs and R. Kunze, 2000 A highly conserved domain of the maize *activator* transposase is involved in dimerization. Plant Cell **12:** 211–224.

Feschotte, C., N. Jiang and S. R. Wessler, 2002 Plant transposable elements: where genetics meets genomics. Nat. Rev. Genet. **3** (5): 329–341.

Gish, W., and D. J. States, 1993 Identification of protein coding regions by database similarity search. Nat. Genet. **3:** 266–272.

Hebsgaard, S. M., P. G. Korning, N. Tolstrup, J. Engelbrecht, P. Rouze *et al.*, 1996 Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. Nucleic Acids Res. **24:** 3439–3452.

Henk, A. D., R. F. Warren and R. W. Innes, 1999 A new *Ac*-like transposon of Arabidopsis is associated with a deletion of the RPS5 disease resistance gene. Genetics **151:** 1581–1589.

Hochstenbach, R., H. Harhangi, K. Schouren and W. Hennig, 1994 Degenerating gypsy retrotransposons in a male fertility gene on the Y chromosome of *Drosophila hydei*. J. Mol. Evol. **39:** 452–465.

Janousek, B., J. Siroky and B. Vyskot, 1996 Epigenetic control of sexual phenotype in a dioecious plant, Melandrium album. Mol. Gen. Genet. **250:** 483–490.

Kunze, R., and C. F. Weil, 2002 The hAT and CACTA superfamilies of plant transposons, pp. 565–610 in *Mobile DNA II*, edited by N. L. Craig, R. Craigie, M. Gellert and A. L. Lambwitz. ASM Press, Washington, DC.

Kunze, R., U. Behrens, U. Courage-Franzkowiak, S. Feldmar, S. Kuhn *et al.*, 1993 Dominant transposition-deficient mutants of maize *Activator* (*Ac*) transposase. Proc. Natl. Acad. Sci. USA **90:** 7094–7098.

Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science **286:** 964–967.

Lebel-Hardenack, S., E. Hauser, T. F. Law, J. Schmid and S. R. Grant, 2002 Mapping of sex determination loci on the white campion (*Silene latifolia*) Y chromosome using amplified fragment length polymorphism. Genetics **160:** 717–725.

Mironov, A. A., J. W. Fickett and M. S. Gelfand, 1999 Frequent alternative splicing of human genes. Genome Res. **9:** 1288–1293.

Moneger, F., 2001 Molecular and evolutionary analysis of a plant Y chromosome. CR Acad. Sci. III **324:** 531–535.

Nakao, S., S. Matsunaga, A. Sakai, T. Kuroiwa and S. Kawano, 2002 RAPD isolation of a Y chromosome specific ORF in a dioecious plant, *Silene latifolia*. Genome **45:** 413–420.

Okada, S., A. W. Kao, B. P. Ceresa, P. Blaikie, B. Margolis *et al.*, 1997 The 66-kDa Shc isoform is a negative regulator of the epidermal growth factor-stimulated mitogen-activated protein kinase pathway. J. Biol. Chem. **272:** 28042–28049.

Parker, J. D., P. S. Rabinovitch and G. C. Burmer, 1991 Targeted gene walking polymerase chain reaction. Nucleic Acids Res. **19:** 3055–3060.

Rosenberg, M., M. Przybylska and D. Straus, 1994 "RFLP subtraction": a method for making libraries of polymorphic markers. Proc. Natl. Acad. Sci. USA **91:** 6113–6117.

Ross, R. J., N. Esposito, X. Y. Shen, S. Von Laue, S. L. Chew *et al.*, 1997 A short isoform of the human growth hormone receptor functions as a dominant negative inhibitor of the full-length receptor and generates large amounts of binding protein. Mol. Endocrinol. **11:** 265–273.

Rubin, E., G. Lithwick and A. A. Levy, 2001 Structure and evolution of the hAT transposon superfamily. Genetics **158:** 949–957.

Sakamoto, K., N. Ohmido, K. Fukui, H. Kamada and S. Satoh, 2000 Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. Plant Mol. Biol. **44:** 723–732.

Shibata, F., M. Hizume and Y. Kuroki, 2000 Differentiation and the polymorphic nature of the Y chromosomes revealed by repetitive sequences in the dioecious plant, *Rumex acetosa*. Chromosome Res. **8:** 229–236.

Srinivasula, S. M., M. Ahmad, Y. Guo, Y. Zhan, Y. Lazebnik *et al.*, 1999 Identification of an endogenous dominant-negative short isoform of caspase-9 that can regulate apoptosis. Cancer Res. **59:** 999–1002.

Steinemann, M., and S. Steinemann, 1991 Preferential Y chromosomal location of *TRIM*, a novel transposable element of *Drosophila miranda*, obscura group. Chromosoma **101:** 169–179.

Steinemann, M., and S. Steinemann, 1992 Degenerating Y chromosome of *Drosophila miranda*: a trap for retrotransposons. Proc. Natl. Acad. Sci. USA **89:** 7591–7595.

Steinemann, M., and S. Steinemann, 1993 A duplication including the Y allele of *Lcp2* and the *TRIM* retrotransposon at the *Lcp* locus on the degenerating neo-Y chromosome of *Drosophila miranda*: molecular structure and mechanisms by which it may have arisen. Genetics **134:** 497–505.

Steinemann, M., and S. Steinemann, 1997 The enigma of Y chromosome degeneration: *TRAM*, a novel retrotransposon is preferentially located on the neo-Y chromosome of *Drosophila miranda*. Genetics **145:** 261–266.

Steinemann, M., and S. Steinemann, 1998 Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda*, a model for sex chromosome evolution. Genetica **103:** 409–420.

Steinemann, M., and S. Steinemann, 2000 Common mechanisms of Y chromosome evolution. Genetica **109:** 105–111.

Steinemann, M., S. Steinemann and F. Lottspeich, 1993 How Y chromosomes become genetically inert. Proc. Natl. Acad. Sci. USA **90:** 5737–5741.

Swofford, D. L., 2002 *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Underwood, A. P., and A. E. Bianco, 1999 Identification of a molecular marker for the Y chromosome of *Brugia malayi*. Mol. Biochem. Parasitol. **99:** 1–10.

Westergaard, M., 1946 Aberrant Y chromosomes and sex expression in *Melandrium album*. Hereditas **32:** 419–443.

Communicating editor: D. Charlesworth