Active miniature transposons from a plant genome and its nonrecombining Y chromosome

Citation for published version:

Digital Object Identifier (DOI):
10.1534/genetics.107.081745

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Genetics

Publisher Rights Statement:
Freely available via Pub Med.

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Active Miniature Transposons From a Plant Genome and Its Nonrecombining Y Chromosome

R. Bergero,1 A. Forrest and D. Charlesworth

Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, Edinburgh, EH9 3JT, United Kingdom

Manuscript received October 18, 2007
Accepted for publication December 16, 2007

ABSTRACT

Mechanisms involved in eroding fitness of evolving Y chromosomes have been the focus of much theoretical and empirical work. Evolving Y chromosomes are expected to accumulate transposable elements (TEs), but it is not known whether such accumulation contributes to their genetic degeneration. Among TEs, miniature inverted-repeat transposable elements are nonautonomous DNA transposons, often inserted in introns and untranslated regions of genes. Thus, if they invade Y-linked genes and selection against their insertion is ineffective, they could contribute to genetic degeneration of evolving Y chromosomes. Here, we examine the population dynamics of active MITEs in the young Y chromosomes of the plant Silene latifolia and compare their distribution with those in recombining genomic regions. To isolate active MITEs, we developed a straightforward approach on the basis of the assumption that recent transposon insertions or excisions create singleton or low-frequency size polymorphisms that can be detected in alleles from natural populations. Transposon display was then used to infer the distribution of MITE insertion frequencies. The overall frequency spectrum showed an excess of singleton and low-frequency insertions, which suggests that these elements are readily removed from recombining chromosomes. In contrast, insertions on the Y chromosomes were present at high frequencies. Their potential contribution to Y degeneration is discussed.

Transposable elements (TEs) have major roles in genome diversification and expansion. Due to their ability to self-replicate, they can proliferate and reach high copy numbers and, if fixed, can be retained in evolutionary lineages across wide taxonomic groups. However, abundance of TEs in eukaryotic genomes varies over several orders of magnitude (Wright and FinneGAN 2001) and the factors that control their population dynamics are not yet completely resolved. TE insertions can be highly deleterious and selective pressures oppose their insertion and thus accumulation. Deleterious effects on fitness derived from TE insertions are of two main types: insertions within or near genic regions can disturb gene functions by changing reading frames or disrupting regulatory motifs (FinneGAN 1992), and chromosomal rearrangements can be caused by ectopic exchange between TE copies at nonhomologous genomic locations (Montgomery et al. 1987; Langley et al. 1988).

In sexually dimorphic organisms whose gender is controlled by sex chromosomes, recombination is suppressed between the Y chromosome, inherited only by male individuals, and the homologous X chromosome (or the Z and W chromosomes in species with female heterogamety). Because Y chromosomes are recombinationally isolated, TE dynamics can be studied in a nonrecombining chromosome in an otherwise recombining background. Y chromosomes should be less affected by deleterious effects of TEs causing chromosomal rearrangements due to ectopic exchanges, because meiotic recombination is suppressed. Thus the main deleterious effects of TEs on fitness of Y chromosomes should be insertions affecting functionally important sequences.

Among TEs, miniature inverted-repeat transposable elements (MITEs) are a class of DNA transposons that move by the trans-activity of a transposase encoded by a related transposable element (Zhang et al. 2001). Unlike other major classes of TEs, MITEs are preferentially located in or near genes (Bureau and Wessler 1994; Feschotte et al. 2002), and, most likely because of their small size (~100–500 bp), insertions often do not cause major disruption of the genes or their regulation (Naito et al. 2006). However, some insertions could be highly deleterious (Yano et al. 2000) and, if they occur within Y-linked genes, they could contribute to genetic degeneration of an evolving Y chromosome.

On Y chromosomes, selection is expected to be ineffective, since recombination is suppressed. Y chromosomes are thus expected to have low effective population size, N_e, due to the “hitchhiking” effects of selection.
[selective sweeps, background selection, and weak Hill–Robertson interference (reviewed by Charlesworth and Charlesworth 2000)]. This expectation is supported by empirical data showing low silent-site diversity of Y-linked genes, compared with their X-linked alleles (Zurovcova and Eanes 1999; Montell et al. 2001). Reduced \( N_e \) should lead to lower efficacy of natural selection, so that mildly deleterious mutations, including MITE insertions, should be able to rise to intermediate frequencies or to fixation (Brookfield and Badge 1997). Moderately deleterious MITE insertions may thus contribute to Y chromosome genetic degeneration.

It is well known that Y chromosomes and neo-Y chromosomes undergo genetic degeneration in the long term (Charlesworth and Charlesworth 2000) and that TEs can quickly accumulate on neo-Y chromosomes (Bachtrog 2003). There are so far no empirical data on the dynamics of MITEs in newly evolving Y chromosomes. We here examine the distribution of MITEs in a dioecious plant species, Silene latifolia, whose sex-chromosome system is not older than 5–10 MY. Silene Y chromosomes are 40% larger than their homologous X chromosomes, but the two sex chromosomes carry a number of homologous genes (Bergero et al. 2007), so that it is unlikely that the larger size of the Y is due to a major autosomal translocation [forming a neo-sex chromosome (Steinemann and Steinemann 1998)]; the most likely repetitive DNA has accumulated. For the few sex-linked gene pairs so far known, S. latifolia Y-linked genes also show lower expression levels than their X counterparts (R. Bergero, unpublished results), suggesting that genetic degeneration is occurring in this species.

To study MITE dynamics, sequences of actively transposing elements are needed. These can be found by scanning complete genome sequences (or sequences of large genome regions) for TE insertions (Duret et al. 2000; Surzycki and Belknap 2000; Rizzon et al. 2002; Wright et al. 2003), but such extensive genomic sequences are difficult to obtain from DNA regions rich in repetitive sequences, such as the Y chromosome (Foote et al. 1992; Holt et al. 2002), or are simply not available from nonmodel species. Furthermore, genome-scan approaches provide no information about the distribution of insertion frequencies, which is needed to test the predictions of the theories outlined above, and it is not always clear whether any given transposable element has recently been active. Indeed, a large fraction of TEs in genomes of higher eukaryotes are probably inactivated copies (fossils) that have lost transposition activity (Smit and Riggs 1996; Feschotte et al. 2002; Pace and Feschotte 2007).

To search for actively transposing MITEs we developed an approach on the basis of assuming that recent transposon insertions or excisions create singleton or low-frequency size polymorphisms, which can be detected in surveys of alleles from natural populations. We used this approach to isolate MITE elements from S. latifolia introns and identified two active subfamilies from this dioecious plant. Transposon display for MITE insertions from both subfamilies was carried out to infer their frequency distributions and compare Y chromosomes with other genome regions sampled from S. latifolia natural populations.

MATERIALS AND METHODS

**Plant material:** One male and one female S. latifolia individual from each of eight European natural populations (supplemental Table 1 at http://www.genetics.org/supplemental/) were used to investigate intron-size polymorphisms. A transposon display was carried out on 48 individuals (24 females and 24 males) derived from a larger set of natural populations (supplemental Table 1) and a collection of 108 F2 plants. The F2 family derived from a single cross between two F1 plants obtained by crossing two parents, one obtained from a French population and one from the Netherlands (Bergero et al. 2007). As these parent plants were not inbred, the elements used as genetic markers are sometimes heterozygous in one parent and sometimes in both (other MITE insertions do not segregate in this family).

Genomic DNA was obtained from fresh leaves using the FastDNA kit (Q-Biogen), following the manufacturer’s instructions.

**Identification of MITEs from intron-size polymorphisms:** Intron variants from a set of 19 genes (two introns were analyzed for each locus, see supplemental Table 2 at http://www.genetics.org/supplemental/) were amplified by PCR and size estimated by standard gel electrophoresis in alleles from natural populations to search for large-size polymorphisms (>150 bp) that could result from recent MITE insertion/excision. Loci were chosen to be single copy or low copy to limit amplification of paralogous genes, which could hinder interpretation of the results. The set also included the Silene sex-linked genes \( 5\text{IX}X/\text{Y}, 6\text{IX}3/\text{Y3}, 6\text{IX}1/\text{Y1}, \text{and} 5\text{ICyP}X/\text{Y} \). Primers were designed on the basis of S. latifolia cDNA sequences. Intron positions were inferred according to gene structures reported for putative Arabidopsis thaliana and Oryza sativa orthologs (Bergero et al. 2007).

PCR products from introns showing size polymorphisms were cloned in a T-tailed pBSK+ vector (Marchuk et al. 1991) and sequenced on an ABI3730 sequencer (Applied Biosystems, Foster City, CA). Alignment of intron-size variants was done using the package Sequencher 4.7 (GeneCodes, Ann Arbor, MI). As MITEs lack transposable coding sequences, other features were used for their identification. These were the presence of terminal inverted repeats (10–15 bp) at the ends of the insertion, target site duplication (TSD) as reported for other MITEs (Wicker et al. 2007), size in the range of 150–500 bp, and extensive secondary structure. The web package MFOLD (Zuker 2003) was used to infer the DNA folding and secondary structure of putative MITE elements.

**MITE insertion variants in a mapping family and in natural populations:** Transposon display (TD) was used to detect segregating MITE elements in an F2 family and MITE polymorphisms and copy numbers in a set of natural populations. TD is an AFLP-based technique (Van den Broeck et al. 1998; Casca et al. 2000) that uses a primer annealing to the adapter and one annealing to conserved regions of the TE element. Although MITEs do not have conserved coding sequences, extensive sequence conservation should occur in members of recently active MITE subfamilies. MITE-specific primers were designed to face outward from and anneal to subterminal sequences of the two MITE elements isolated from S. latifolia.
(SlTo1 and EITRI). The procedures were as outlined in Casa et al. (2004) with the following modifications. Genomic DNA (0.8–1.5 µg) was digested with DpnII for 3 hr at 37°C. After inactivation of DpnII by incubation at 65°C for 10 min, restriction fragments were ligated to a DpnII linker. The ligation reactions containing 200 units T4 DNA ligase (New England Biolabs, Ipswich, MA), 20 units BamIII, and 12 µM of DpnII adapter were performed for 12 cycles, each consisting of 30 min at 16°C and 10 min at 37°C. The DpnII adapter was obtained by spontaneous annealing at room temperature of the oligonucleotides TDA1A1 (5’-GACAGTTGTGTAACCTCGGATG-3’) and TDA2A1 (5’-GATCTCCAGGATGTAACAGCTG-3’). Adapter dimer formation in the ligation reaction (which could significantly decrease the availability of adapters for ligation with genomic DNA) was avoided by designing a 5’-GA overhang at one end of the adapter and by forming an ex novo BamHI site at the other end when adapter dimers formed. Dimers were destroyed by adding BamHI to the ligation reaction.

The DpnII library was amplified by a round PCR of 20 cycles using adapter-specific primers (with one specific base) and the MITE-specific external primer (EITRIext, 5’-TAA TAACGTTGTCCTCGCC-3’ and SlTo1ext, 5’-TCCATTCCA ATCTCATTCAAGAG-3’). Thermocycling conditions were as follows: 4 min at 94°C; 20 cycles at 30 sec at 95°C, 40 sec at 50°C, 30 sec at 72°C; and 5 min at 72°C. One microliter of the PCR reaction was used as template in a nested touchdown PCR (DpnII/C176 and 10 min at 37°C). The procedures were as outlined in C EITRIII library was amplified by a first-round PCR of 20 cycles using adapter-specific primers (with one specific base and the MITE-specific internal primer (EITRIint, 5’-TCC GTGTCCTAAATTTCTG-3’ and SlTo1int, 5’-GAGAGCAAAC CAAACCCCC-3’), with the following thermocycling conditions: 5 min at 95°C, followed by 10 annealing cycles at 0.5°C decreasing and 18 cycles at 50°C. The selective amplification was carried out using a hot-start TAQ polymerase (JumpStart; Sigma, St. Louis), which was found to increase product yields, especially for bigger-size bands (>300 bp), and to produce better electropherograms.

VIC-labeled PCR products were first diluted (1:25) in distilled water, further diluted (1:10) in formamide containing the size standard GeneScan 500-LIZ (Applied Biosystems), and directly separated by capillary electrophoresis on an ABI-3730 DNA analyzer (Applied Biosystems). Transposon insertions were scored using the software Genemapper v. 3.7 (Applied Biosystems), and their frequencies were analyzed and plotted using the R statistics package (http://www.r-project.org).

MITE sequences from this study were deposited in the GenBank databases (accession nos. EU334132 and EU354133).

Data analyses: The average numbers of MITE insertions per haploid genome were estimated from electropherograms obtained from the set of 24 females (thus excluding Y-linked TE insertions whose properties may be unusual). Because of a preponderance of rare insertions, an estimate of insertion frequencies can be obtained from the observed frequencies of the recessive genotypes (nulls) of each insertion. To obtain the frequency, \( p_i \), of insertions at the \( i \)th site, we assumed that the population is in Hardy–Weinberg equilibrium, since the species is outcrossing, and used \( p_i = 1 - q_i^2 \), where \( q_i \) is the frequency of the recessive genotype at this site (with no band on the gel). The expected number of TE insertions per haploid genome is the total number of insertions weighted by their frequencies (\( N_m = \sum_{i=1}^{n} N_{mi} p_i \)).

Indels in the intervening sequence between a MITE insertion and the restriction site will create a new “spurious” polymorphic TE insertion. We estimated the proportion of the observed polymorphisms due to such events by examining segregation of 203 MITE insertions in an F2 progeny made by crossing two F1 individuals from a cross between two outbred natural populations. Pairs of transposon display bands that segregated as alternatives in repulsion were counted as probable insertions that are descended from a single ancestral insertion, but that have undergone indel events since the insertion.

### RESULTS

**Intron insertion polymorphisms due to MITE activity:** A total of 38 primer pairs amplifying intronic regions from 19 loci were used to search for large insertion polymorphisms in European populations of *S. latifolia*. Sixteen plants were included in the survey, and 5 of the 38 introns showed size polymorphisms, with size differences >150 bp (Table 1).

<table>
<thead>
<tr>
<th>Locus name</th>
<th>Genomic location</th>
<th>Intron insertion size (bp)</th>
<th>% AT</th>
<th>MITE designation</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIAATPS6</td>
<td>Autosomal</td>
<td>2</td>
<td>232</td>
<td>57</td>
</tr>
<tr>
<td>SIAnk</td>
<td>Autosomal</td>
<td>5</td>
<td>407</td>
<td>51</td>
</tr>
<tr>
<td>SIGypY</td>
<td>Y-linked</td>
<td>2</td>
<td>290</td>
<td>58</td>
</tr>
<tr>
<td>SIPi</td>
<td>Autosomal</td>
<td>5</td>
<td>348</td>
<td>65</td>
</tr>
<tr>
<td>SIX3</td>
<td>X-linked</td>
<td>10</td>
<td>226</td>
<td>63</td>
</tr>
</tbody>
</table>

PCR amplification of intron 5 from a *S. latifolia* gene (SIAnk) produced an ~600-bp ampiclon in all individuals, but a single male plant appeared to be heterozygous, having an additional larger PCR product (~1000 bp, Figure 1). A single 407-bp insertion delimited by 11-bp terminal inverted repeats (with the sequence 5’-CTAGGTAGCAC-3’) and 8-bp TSDs was confirmed in the larger amplicon by sequencing. The TSD has the imperfect palindromic sequence 5’-CTCTTGGAG-3’. Excluding the insertion, the 1-kb product differed from its allelic counterpart in the same plant by one base substitution and a 3-bp indel. These data suggest that the long and short sequences are allelic. The presence of terminal inverted repeats (TIRs), TSDs, and extensive secondary structure strongly suggests that this insertion is a MITE element. Classification of nonautonomous TEs relies on the TIR motifs and TSD sizes (Wicker et al. 2007). The size of the TSD (8 bp) suggests that this element is either a hAT or a P element, but the TIR motifs known from these two classes of elements were not found. We therefore identify it as the first MITE obtained from our study species, and, given its uncertain classification, we name it EITRI.

In contrast to this singleton polymorphic insertion, a size polymorphism due to a singleton excision was observed in a sex-linked gene. PCR amplification of intron 2 of the recently described *S. latifolia* sex-linked gene SIGypX/Y (Bergero et al. 2007) produced two amplicons (710 and 1000 bp). Segregation analysis of these bands clearly showed Y linkage of the longer intron variant.
Sequence alignment of these two variants shows an insertion of 290 bp, delimited by a 14-bp inverted repeat, and the inserted sequence exhibits the potential for extensive secondary structure (Figure 2). Surprisingly, its TIR (5’-GGGGGTGTTTGGTT-3’) matches perfectly the TIR region of a Tourist element (Zm20) isolated from Zea mays (Bureau and Wessler 1994) and TIR consensus sequence from 21 rice PIF families (Zhang et al. 2004), probably because of high conservation of the catalytic domain of the transposase. Furthermore, a 3-bp TTA motif flanked this insertion, which is typical for TSDs of Tourist (mPIF) elements (Jurka and Kapitonov 2001). We therefore classified this as a Tourist-like element and named it SlTo1. In a sample of eight Y chromosomes from natural populations, we found a singleton excision with a clearly visible footprint in the SlCypY sequence alignment (Figure 2). A search for this insertion in Cyp orthologs of other, closely related, dioecious Silene species revealed the same transposable element inserted in the Y chromosomes of S. diclinis and S. dioica; the sequence identity of these MITE insertions was estimated to be 99%. Although there are reported cases of independent insertions in the identical site (Walker et al. 1997), the most parsimonious explanation for this MITE in the SlCypY gene is that these three sister species split after X-Y recombination stopped in this region and that this Tourist insertion occurred before this time. This is consistent with the fact that the divergence between the SlCypX and the Y copies ($K_s = 6.1\%$, from Bergero et al. 2007) is considerably higher than that between these species ($K_s = 4.4\%$).
We detected three further intron-size polymorphisms due to insertions of unknown origins (Table 1). All the polymorphic insertions from these three loci appear to be at low to intermediate frequencies, but none are singleton insertions or excisions. These insertions did not have inverted repeats, nor did analysis of their sequences suggest extensive secondary structure. Thus they are not recognizable MITEs. The origin of these large insertion variants is puzzling. Their intermediate frequencies in natural populations suggest that these are not recent insertions, and this is consistent with the absence of the conserved features of MITE sequences. They could represent relics of MITEs, deleted for part(s) of the sequence, including the TIRs [solo LTRs are known in other plant genomes and are caused by deletions (Devos et al. 2002; Ma et al. 2004)], or the TIR sequences may have become unrecognizable due to mutations.

Transposon display of genomic and Y-linked MITE insertions: A TD analysis of the two MITE elements, EITRI and SITo1, was carried out on a set of 48 individuals collected from 24 European natural populations (supplemental Table 1). Using the frequencies of null alleles (see MATERIALS AND METHODS), we estimated an average of 230 copies of EITRI and 130 copies of the SITo1 element per haploid genome.

Segregation of MITE insertions in a full-sib F2 family allowed us to recognize MITE insertions in the Y chromosome. Twenty-three EITRI elements and 16 SITo1 elements showed a clear segregation pattern of complete Y linkage. From the estimated average number of MITE insertions per megabase in the genome as a whole, we computed the predicted number of Y-linked insertions that should be found on a Y chromosome. Taking into account the physical size of the Y chromosome [the largest S. latifolia chromosome, estimated to be 570 Mb (Siroky et al. 2001)], and assuming a uniform distribution of TE insertions in the S. latifolia genome, there are significantly fewer MITE insertions in the Y chromosome than the expected copy numbers for both the EITRI and the SITo1 subfamilies (49 and 28, respectively; \( \chi^2 = 18.9, P < 0.0001 \)).

An excess of singleton or low-frequency MITE genomic insertions was observed from the TD analysis of the 24 female plants sampled from natural populations (Figure 3); only a small fraction (5%) of insertion sites were at medium or high frequencies \( (p_i > 0.3) \). In contrast, the frequency spectrum for Y-linked insertions was markedly shifted toward high frequencies, with a remarkable paucity of singleton and low-frequency insertions (Figure 3). Thus, the ratio of fixed to polymorphic Y-linked insertions was greater than the ratio of fixed to polymorphic genomic insertions, due to an excess of fixed Y-linked insertions \( (3/23 \text{ vs. } 1/1923) \), for the Y-linked and the other insertions, respectively. A \( \chi^2 \)-test for independence, with Yates’ correction for continuity, showed that the two ratios were highly significantly different \( (\chi^2 = 113.99, P < 0.000001) \).

**DISCUSSION**

MITEs and their contribution to the genetic degeneration of evolving Y chromosomes: With their ability to invade genic regions, MITEs could contribute to the genetic degeneration of evolving Y chromosomes, if selection against their removal is ineffective. Here we show that MITE elements from two active subfamilies are invading the evolving S. latifolia Y chromosome and are present at intermediate and high frequencies or fixed on this chromosome. This strongly contrasts with the preponderance of singleton and rare insertions in the rest of the genome (insertions with frequencies \( < 0.1 \) account for 78% of all observed insertions), which suggests that these elements are readily removed from recombining chromosomes. MITE insertions in Y-linked genes at intermediate to high frequencies could change gene expression and lower functions of Y-linked genes. Future empirical work should test this possibility directly, for example, by testing whether the presence of insertions correlates with changed expression of Y-linked genes.
A caveat to estimating numbers of polymorphic TE insertions from transposon display is that a proportion of the polymorphisms may not be due to insertion/excision events, but to indel mutations and/or mutations changing restriction sites. Transposon display cannot distinguish bona fide polymorphic TE insertions from polymorphisms created by indels that change the size of TD fragments or SNPs that modify restriction sites. This will lead to overestimated numbers of polymorphic TE insertions, which will probably affect autosomes and X chromosomes more than Y chromosomes because sequencing studies show that Y-linked sequences have lower diversity than is found in other genome regions (Filatov et al. 2000, 2001; Filatov 2005; Laporte et al. 2005). The relative number of low-frequency polymorphic insertions may thus be somewhat inflated for the genome regions other than the Y. However, this effect is probably very slight. Although indels are common in intronic sequences from this plant species (Bergero et al. 2007), indels postdating MITE insertion events (see Materials and Methods for how these are ascertained) are seen in only a small proportion (2%) of 203 MITE insertions segregating in an F2 family. This supports our assumption of a recent age of most of the MITE insertions observed in our data set and excludes the possibility that intervening indels have substantially inflated our estimates of MITE polymorphisms. Similarly, destruction of restriction sites, or creation of new ones in the region between the original restriction sites and the TE insertion, should be infrequent in the short period after a TE insertion. Using the approach of Net and Li (1979), the probability of a changed restriction site for a 4-base cutter (DpnII in our study) by the time t after a TE insertion, given a rate λ of nucleotide substitutions per unit time, is (1 − e^{-λt}), and the expected probability of a new restriction site appearing in the intervening region (based on surveyed fragment sizes of 175 ± 78 bp) is 0.68(1 − e^{-λt}). If MITE insertions are recent, t is small, and these probabilities become very small.

MITES and the dynamics in an asexual (nonrecombining) genetic background: High MITE insertion frequencies will lead to an increased mean copy number of insertions per Y chromosome. The high frequencies of MITE insertions in the S. latifolia Y chromosome indicate increased chances of finding a TE per site. It is of interest also to ask whether there is also a higher total number of Y-linked insertions than the expected average number for a genomic region of similar size (i.e., whether there are more insertion sites per megabase in the Y than in other genome regions, which might indicate loss of functional sequences and therefore genetic degeneration). Computing the expected number of Y-linked MITEs from the estimated average number of insertions per megabase, Y-linked MITEs appear to be somewhat underrepresented (23 ETRI and 16 STO1 insertions observed, vs. expected numbers of 49 and 28 insertions, respectively). The estimated numbers of TE copies were estimated from a single Y chromosome (the one in our family), and the numbers are likely to vary among Y haplotypes. However, given that most of the Y-linked TEs are fixed or at high frequencies (see Figure 3 above), any such variability should be minor and will not greatly affect our conclusions.

Our result contrasts with findings from the neo-Y of Drosophila miranda. In a survey of 12 D. miranda lines, Bachtrog (2003) estimated that the Y-linked retrotransposon insertions were fixed and thus inferred that the number of insertions present in the neo-Y exceeds that in the homologous regions of the neo-X chromosome. However, unlike retrotransposons (which use a copy-and-paste mechanism), MITEs are nonautonomous versions of DNA transposons, which use a cut-and-paste mechanism for their movement and replication (Engels et al. 1990). It is therefore plausible that their accumulation will differ from that of retrotransposons. Another reason for underrepresentation of Y-linked MITEs is that they may be preferentially located either in regions with high-recombination rates [as reported for DNA transposons in the Caenorhabditis elegans genome (Duret et al. 2000)] or in regions with high gene density (Wright et al. 2003). Mammalian Drosophila Y chromosomes are notoriously low in gene content (Lahn et al. 2001), and Silene Y chromosomes could be in a stage where degeneration is already eroding gene content.

Nevertheless, the Silene Y chromosome has a DNA content 40% larger than the homologous X chromosome (Siroky et al. 2001), and it seems likely that accumulation of TEs could explain this size difference. Different behavior of different elements has been found on the nonrecombining fourth chromosome of D. melanogaster; there is a significant accumulation of non-LTR retrotransposon elements, but not of the much younger class of LTR retrotransposons (Kaminsker et al. 2002; Bergman and Benssos 2007). To test whether TE classes other than the MITEs studied here have accumulated, more TE types should be studied in the future, especially RNA transposons.

The role of MITEs in plant genome diversification and expansion: Our approach of screening for large polymorphic insertions within a small set of intronic sequences revealed active MITE elements in the nonmodel plant S. latifolia, producing either singleton insertions or excisions. The application of this approach in nonmodel species is thus straightforward and will greatly facilitate discovery of new active MITE families, to study their accumulation and locations in eukaryotic genomes.

The finding of hundreds of MITE copies per haploid genome suggests that the S. latifolia genome probably contains many active MITEs. A large fraction of the genome of many higher plants is repetitive, and MITEs are reported to represent the largest component among O. sativa transposable elements and a major fraction of...
of the Arabidopsis genome (Jiang et al. 2004). In the genus Silene, estimated genome sizes range from 1100 Mb in S. vulgaris to 26-46 Mb in S. latifolia and to 3500 in S. chalcedonica (Storky et al. 2001; Meagher et al. 2005). These species all have the same chromosome number \(2n = 24\), like all nonpolyploid species in the genus for which data exist (Goldblatt 1981), which excludes a genome duplication as the cause of genome size differences. This high variability in genome sizes suggests that differential accumulation of repetitive sequences and MITEs could have contributed to genome size expansion over relatively short evolutionary times (the genus Silene is not older than a few million years, since the largest synonymous-site divergence between any pair of species in the genus is 20.8%, based on 1791 bp of concatenated coding sequences from five loci).

This research was supported by the National Environmental Research Council (grant no. NE/B504230/1).

LITERATURE CITED


Communicating editor: D. Voytas