Rapid Degeneration of Noncoding DNA Regions Surrounding SlAP3X/Y After Recombination Suppression in the Dioecious Plant Silene latifolia

Kotaro Ishii,* Rie Nishiyama,† Fukashi Shibata,‡ Yusuke Kazama,§ Tomoko Abe,*§ and Shigeyuki Kawano,**†
*RIKEN Nishina Center, Wako, Saitama 351-0198, Japan, †RIKEN Center for Sustainable Resource Science, Yokohama, Kanagawa 230-0045, Japan, ‡Institute of Plant Science and Resources, Okayama University, Kurashiki 710-0046, Japan, §RIKEN Innovation Center, Wako, Saitama 351-0198, Japan, and **Department of Integrated Biosciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba 277-8562, Japan

ABSTRACT  Silene latifolia is a dioecious plant with heteromorphic XY sex chromosomes. Previous studies of sex chromosome–linked genes have suggested a gradual divergence between the X-linked and the Y-linked genes in proportion to the distance from the pseudoautosomal region. However, such a comparison has yet to be made for the noncoding regions. To better characterize the nonrecombining region of the X and Y chromosomes, we sequenced bacterial artificial chromosome clones containing the sex chromosome–linked paralogs SlAP3X and SlAP3Y, including 115 kb and 73 kb of sequences, respectively, flanking these genes. The synonymous nucleotide divergence between SlAP3X and SlAP3Y indicated that recombination stopped approximately 3.4 million years ago. Sequence homology analysis revealed the presence of six long terminal repeat retrotransposon-like elements. Using the nucleotide divergence calculated between left and right long terminal repeat sequences, insertion dates were estimated to be 0.083–1.6 million years ago, implying that all elements detected were inserted after recombination stopped. A reciprocal sequence homology search facilitated the identification of four homologous noncoding DNA regions between the X and Y chromosomes, spanning 6.7% and 10.6% of the X chromosome–derived and Y chromosome–derived sequences, respectively, investigated. Genomic Southern blotting and fluorescence in situ hybridization showed that the noncoding DNA flanking SlAP3X/Y has homology to many regions throughout the genome, regardless of whether they were homologous between the X and Y chromosomes. This finding suggests that most noncoding DNA regions rapidly lose their counterparts because of the introduction of transposable elements and indels (insertion–deletions) after recombination has stopped.

KEYWORDS  SlAP3 transposon sequence divergence sex chromosome Silene latifolia

Many plants, including the model plant Arabidopsis thaliana, are bisexual. However, approximately 6% of angiosperms are dioecious, meaning that male plants produce only male flowers and female plants produce only female flowers (Renner and Ricklefs 1995). Some dioecious plants have sex chromosomes with varying sizes and forms. Carica papaya (60 Mb) (Ming et al. 2008) and Asparagus officinalis (65 Mb) (Telgmann-Rauber et al. 2007), for example, have homomorphic sex chromosomes. In contrast, Marchantia polymorpha (X: 20 Mb; Y: 10 Mb) (Yamato et al. 2007) and Rumex acetosa (X: 740 Mb; Y1: 520 Mb; Y2: 410 Mb) (Mosiolczek et al. 2005) have heteromorphic sex chromosomes. In Silene latifolia, sex is determined by heteromorphic sex chromosomes, with a pair of X chromosomes in females and both an X chromosome and a Y chromosome in males. The Y chromosome is estimated to be 570 Mb (Široký et al. 2001; Liu et al. 2004), representing approximately 9% of the total genome (Matsunaga et al. 1994), and is 1.4-times larger than the X chromosome (Matsunaga et al. 1994). Thus, the X and Y chromosomes of S. latifolia provide a good opportunity to study the evolutionary history of heteromorphic sex chromosomes during plant evolution.
Sex chromosomes have evolved independently in many plant groups (Charlesworth 2002). Sex chromosomes are thought to have originated from a pair of autosomes with two sexually antagonistic mutations in plants and in animals. It is also assumed that the chromosomal region harboring these mutations was subject to selection-driven suppression of recombination (Charlesworth 2013), which may have been facilitated by chromosomal inversions (Lemaître et al. 2009; Wang et al. 2012) and translocations (Charlesworth and Charlesworth 1980). In mammals, such nonrecombining genomic regions are known to have expanded over evolutionary time (Iwase et al. 2003).

One way to estimate the age of sex chromosomes is to study X-Y divergence (Charlesworth 2013). Estimating the date of sex chromosome emergence is possible when using the sex-determining gene itself and the use of silent site divergence of an X-Y gene pair allows for estimation of the time when recombination stopped between the two chromosomal regions in which the gene pair resides. Using several X-Y gene pairs, dates associated with recombination suppression have been estimated in species of mammals (Lahn and Page 1999), birds (Lawson-Handley et al. 2004; Nam and Ellegren 2008), and plants [S. latifolia (Bergero et al. 2007) and C. papaya (Wang et al. 2012)], revealing differences in the estimated dates based on the chromosomal regions in which the pairs were located. Chromosomal regions exhibiting different silent site sequence divergence estimates are referred to as “evolutionary strata” (Lahn and Page 1999). The oldest stratum in S. latifolia is thought to have appeared 5–10 million years ago (MYA) (Nicolas et al. 2005). By mapping eight X-linked genes, it has been reported that silent site divergence between X-Y gene pairs increased in proportion to the distance of the gene pair from the pseudoautosomal region (PAR), and it has been suggested that recombination between the X and Y chromosomes stopped in progressive steps that formed two evolutionary strata (Bergero et al. 2007). Bergero et al. (2013) showed another stratum that was formed by additions of genome regions. However, previous studies focused on the comparison of only X-Y gene pairs or the accumulation of repetitive sequences at the chromosomal level (Kejnovsky et al. 2009). Thus, the nonrecombining regions that comprise each evolutionary stratum remain uncharacterized.

The S. latifolia MADS box gene, SlAP3, which exhibits significant similarity to A. thaliana APETALA3, was first isolated by Matsunaga et al. (2003) and was found to occur in two copies, SlAP3Y and SlAP3X. SlAP3Y has only been detected in male plants when genomic PCR is performed using SlAP3Y-specific primers and is thus known to be located on the Y chromosome. In contrast, SlAP3A was thought to be located on the autosome, because this gene was first detected by genomic PCR using SlAP3A-specific primers and flow-sorted X chromosomes and autosomes (Matsunaga et al. 2003). It was also later amplified by Cegan et al. (2010) by using microdissected X chromosomes, but not microdissected autosomes; the gene has since been renamed SlAP3X. Reaching the same conclusion, Nishiyama et al. (2010) conducted a segregation analysis of SlAP3X in which different lines of S. latifolia were crossed and the segregation of the male parent–derived sequence was examined. The initial misunderstanding of SlAP3X mainly results from the use of a localization test of flow-sorted chromosomes derived from cultured root cells in which the translocation including SlAP3X would have occurred. Cegan et al. (2010) compared promoter sequences of SlAP3X and SlAP3Y and revealed that a specific sequence was inserted in the SlAP3Y promoter. This inserted sequence shows some homology to the promoter sequence of MROS1, which is specifically expressed in male plants (Matsunaga et al. 1996, 1997). The divergence of noncoding regions is possibly a driving force behind the degeneration of X and Y chromosomes, resulting in differential expression of X-linked and Y-linked genes.

Based on the synonymous nucleotide divergence calculated in the coding regions of SlAP3X and SlAP3Y, we estimated that recombination between the two genes stopped approximately 3.4 MYA. Comparison of intron sequences within the pair of genes revealed that homologous regions have been segmentalized by insertions and deletions (indels) and contain repetitive sequences. Additionally, we compared sequences surrounding the gene pair; predictably, the transposons classes observed in these regions differed between the X and Y chromosomes. We estimated insertion times of predicted long terminal repeat (LTR) retroelements based on the nucleotide divergence between left and right LTRs, which implies that they had likely been inserted in these regions after recombination between SlAP3X and SlAP3Y stopped. These data provide the first evidence related to nucleotide resolution of sequence divergence between a pair of noncoding DNA regions on the sex chromosomes of S. latifolia after recombination stopped.

MATERIALS AND METHODS

Plant materials

An inbred S. latifolia line, the K line (Kazama et al. 2003), was used for all experiments. A second line, the B line, was used for linkage mapping of the X chromosome. The original plants of K and B lines were provided by the University of Oslo Botanical Garden, Norway (K line), and the Royal Botanic Gardens, Kew, United Kingdom (B line). F1 progeny of the two phyletic lines were obtained by crossing a male plant of the K line with a female plant of the B line. F2 progeny were generated by crossing an F1 male with an F1 female. Plants were grown in pots in a regulated chamber at 23°C under a 16:8-hr light/dark cycle. Young leaves were used for genomic DNA isolation.

Bacterial artificial chromosome sequencing

The clone 2b5E (renamed as 13d11Eb) containing SlAP3X was isolated from a male S. latifolia bacterial artificial chromosome (BAC) library, as described by Ishii et al. (2010), using the 4D-PCR method (Asakawa et al. 1997; Ishii et al. 2008). All BAC clones (13d11Ea, 13d11Eb, and 7a8D) were purified using the PowerPrep HP Plasmid Purification Kit from OriGene Technologies (Rockville, MD) and were further purified by CsCl gradient ultracentrifugation. Emulsion PCR and 454 sequencing were performed on the Roche 454 FLX platform according to manufacturer’s instructions. Sequence assembly was performed using the GS De novo Assembler. A summary of the sequencing data is shown in Supporting Information, Table S1. Contigs consisting of more than 20 reads were first extracted and used to construct supercontigs in Sequencher (Gene Codes, Ann Arbor, MI). Contigs of 13d11Ea and 13d11Eb were analyzed together. All gaps between the contigs were PCR-amplified and sequenced by the Sanger sequencing method.

Sequence analysis

Alignments of intron sequences of SlAP3X and SlAP3Y were conducted using the Needle program within the EMBOSS 6.5 package (Rice et al. 2000) with default settings, except that the gap penalty was set to 30.0 and the extend penalty was set to 0.0. Dot plot analyses were performed using HarrPlot 3.1.1 included in GENETYX version 11 with the unit size to compare set to 10 and the dot plot matching number set to 8. Homology searches were performed with the nucleotide sequence database of the National Center for Biotechnology Information using the TBLASTX program and with the repetitive sequence database, Repbase (Jurka et al. 2005) of the Genetic
Information Research Institute (taxon: Viridiplantae), using the BLASTN program. These programs are contained in the BLAST+ package (Camacho et al. 2009). Repetitive sequences, including LTRs, simple repeats, and other repeat sequences, were detected by homology searches of the 13d11E or 7a8D sequences with their own sequences using the BLASTN program. The E-value threshold was set to 1E−10. Homologous regions were detected from homology searches of the 13d11E or 7a8D sequences using the BLASTN program. Thresholds for hit length, E-value, and identity were set to 100 bp, 1E−50, and 80%, respectively. The pair of coding regions of SlAP3X and SlAP3Y, the pairs of LTRs, and the pairs of homologous regions between the 13d11E and 7a8D sequences were aligned using the program MUSCLE (Edgar 2004). The synonymous nucleotide distance (k) between the coding regions of SlAP3X and SlAP3Y was estimated using the Kumar method (Nei and Kumar 2000). The nucleotide distances (k) of the pairs of the LTRs and the pairs of the homologous regions were estimated using the Tamura-Nei model (Tamura and Nei 1993) within MEGA5 (Tamura et al. 2011). An average substitution rate (r) of 1.8 x 10−8 substitutions per synonymous site per year (Du et al. 2010) was used for calculations. The time (T) since recombination stopped between SlAP3X and SlAP3Y was estimated using the following formula: T = k/2r.

**Genomic Southern blot analysis**

Southern blot analysis methods were modified from those described previously (Kazama et al. 2006). Genomic DNA was extracted from S. latifolia leaves using a Nucleo PrepPure Genomic DNA Extraction Kit (GE Healthcare, Little Chalfont, England). Genomic DNA (15 μg) was digested with EcoRI for 12 hr. The concentrations of the digests were measured, and the digests were loaded onto a 1.0% (w/v) agarose gel and then transferred to an Immobilon-Ny+ membrane (Merck, Darmstadt, Germany). Hybridization and signal detection were performed using the AlkPhos Direct Labeling and Detection System (GE Healthcare), and each probe was amplified individually by PCR with locus-specific primers (Table S2). The hybridized membranes were visualized using CDP-Star (GE Healthcare) at room temperature with appropriate exposure times.

**FISH analysis**

Fluorescent in situ hybridization (FISH) was performed as previously described (Ishii et al. 2010), with minor modifications. Two probes for the homologous (I) and nonhomologous (a) regions were prepared from the same PCR products produced in the genomic Southern blot analysis using DIG-Nick and the Biotin-Nick Translation Mix (Roche Diagnostics, Basel, Switzerland), respectively. Chromosomal DNA was denatured at 70°C for 1 min in 2x SSC buffer containing 70% formamide. The chromosome preparations were dehydrated immediately by 5-min treatments with 70% ethanol at −20°C and 100% ethanol at room temperature. The preparations were then dried for 30 min at room temperature. Each slide was loaded with 10 μL of the hybridization mixture containing 12.5 ng probe, 50% formamide, 10% dextran sulfate, and 2x SSC buffer. The slides were washed twice in 2x SSC buffer at 37°C. The signals were fluoresceinlabeled with antidigoxigenin-rhodamine (Roche Diagnostics) and avidin-Alexa Fluor 488 (Molecular Probes, Eugene, Oregon) for the digoxigenin-labeled probe and biotin-labeled probe, respectively. Preparations were counterstained with 4’,6-diamidino-2-phenylindole (DAPI). The fluorescent Alexa Fluor 488, DAPI, and rhodamine images were detected consecutively under a Leica Q550 cytogenetic workstation (Leica Microsystems, Wetzlar, Germany) equipped with a black-and-white charged-coupled device camera (CoolSNAP HQ; Nippon Roper, Tokyo, Japan) using the SpectraVision filters SpectrumGreen, SpectrumDAPI, and SpectrumOrange, respectively (Vysis, Chicago, IL).

**X chromosome linkage mapping**

To genotype the population, genomic PCRs were performed using K line-specific primers for SlAP3X, SlX1, and SlX4 (Table S3) on 96 F2 progeny derived from a cross between a K line male and a B line female. Restriction fragment length polymorphism analysis was used to genotype the DD44X locus. Genomic DNA fragments were amplified with DD44X-specific primers (Table S3) from the 96 F2 progeny, digested with NcoI, and analyzed by electrophoresis. The Kosambi mapping function was used to compute genetic distances in cM (Kosambi 1944).

**RESULTS**

**Structure of the SlAP3X/Y transcriptional region**

The number of synonymous nucleotide divergences per synonymous sites between translated regions of SlAP3X and SlAP3Y was 0.12 (SE, 0.034). Nicolas et al. (2005) estimated that in the oldest evolutionary strata, recombination stopped 5−10 MYA by using the average synonymous substitution rates of two genes (1.8 x 10−8 substitutions per synonymous site per year) in the Brassicaceae species (Koch et al. 2000). Similarly, we estimated that recombination between SlAP3X and SlAP3Y stopped 3.4 MYA (SE, 0.94).

SlAP3X has five introns and SlAP3Y has six introns (Nishiyama et al. 2010). Exon V and exon VI of SlAP3Y are homologous to the anterior half and the posterior half of exon V of SlAP3X (Nishiyama et al. 2010), respectively, which indicated that intron 5 of SlAP3Y was inserted after recombination stopped. Thus, the five pairs of introns shown in Table 1 were considered to be homologous because they underwent recombination before recombination suppression. Nishiyama et al. (2010) previously reported that intron 2 of SlAP3X was 561 bp long, whereas intron 2 of SlAP3Y was 24.4 kb long, containing two retroelements and one telomere-like sequence. Therefore, we concluded that the majority of intron 2 of SlAP3Y was inserted after recombination stopped. We aligned each pair of introns and calculated the similarities. Although lengths of the paired introns differed between SlAP3X and SlAP3Y, calculated sequence similarities (excluding gaps) were between 76% and 82% (Table 1).

We conducted dot plot analyses of each pair of introns (Figure 1). All five pairs contained regions of sequence homology, which implied that the pairs were homologous before recombination suppression. However, these homologous regions did not span the full lengths of the introns but were interspersed like island chains. To investigate whether introns consisted of unique or repetitive sequences, we designed probes inside the introns and performed genomic Southern blot analyses (Figure 2). The probes designed for the introns showed three patterns: a strong smear pattern (introns 1 and 2 of SlAP3X and introns 1 and 4 of SlAP3Y); a weak smear pattern (intron 3, 4, and 5 of SlAP3X and introns 3 of SlAP3Y); and a multiple-banded pattern (intron 5 and 6 of SlAP3Y). From this, it was revealed that all introns contained repetitive sequences; the distinctness of these patterns implied that each intron contained a different type of repeat sequence.

**Transposons in the peripheral regions of SlAP3X/Y**

To obtain sequences in the peripheral regions of SlAP3X and SlAP3Y, we sequenced the following three BAC clones using the 454-FLX sequencing platform: 13d11Ea containing SlAP3X; 7a8D containing SlAP3Y (Nishiyama et al. 2010); and 2b5E containing SlAP3X (isolated in this study). The sequences of 13d11Ea and 2b5E spanned the
same chromosomal region, which indicated that they were derived from the same genome fragment; 2b5E was thus renamed 13d11Eb. After the vector sequence was removed, we obtained a 115,665-bp sequence flanking SIAP3X (13d11E) and a 72,683-bp sequence flanking SIAP3Y (7a8D) (Figure 3).

We conducted TBLASTX searches of the National Center for Biotechnology Information nucleotide database using the 13d11E and 7a8D sequences as the query sequences (Table S4 and Table S5). In 13d11E, 11 regions were homologous to the coding regions of LTR retrotransposons and one region was homologous to a coding region of a DNA transposon. In 7a8D, six regions were homologous to coding regions of LTR retrotransposons and two regions were homologous to LINEs. BLASTN searches of the Genetic Information Research Institute Repbase database (taxon: Viridiplantae) using 13d11E and 7a8D as query sequences revealed that 22 regions in 13d11E were homologous to LTR retrotransposons, and seven regions in 7a8D were homologous to LTR retrotransposon–like elements were identified with nested structures (Figure 4). The BLASTN searches also revealed the existence of two simple repeat sequences in 13d11E and seven simple repeat sequences in 7a8D. Six out of seven simple repeat sequences in 7a8D are found in one LTR retrotransposon–like element. The majority of sequences of the three LTR retrotransposon–like elements in 13d11E were homologous to one another; similar results were observed for two DNA transposon–like elements in 13d11E, which indicated that they were genetically related to each other.

Transposons inserted after recombination suppression

Du et al. (2010) estimated the insertion dates of LTR retrotransposons using genetic distances calculated between left and right LTR sequences. In the same manner, we calculated the genetic distances between the left and right LTR sequences of the LTR retrotransposon–like elements detected in 13d11E (SIAP3X) and 7a8D (SIAP3Y), from which we estimated the dates when they were inserted (Table 2 and Figure 5). All the investigated LTR retrotransposon–like elements were estimated to be inserted after recombination between SIAP3X and SIAP3Y had stopped. Even the oldest one, which is the copia retrotransposon–like element, was presumed to be inserted 1.6 MYA.

| Table 1 Homology between introns of SIAP3X and SIAP3Y |
|-----------------------------------------------|
| **SIAP3X (13d11E)** | **SIAP3Y (7a8D)** |
| **Intron** | **Start, bp** | **End, bp** | **Length, bp** | **Start, bp** | **End, bp** | **Length, bp** | **Identity (%)** |
| 1 | 86,023 | 86,127 | 105 | 1 | 34,286 | 34,402 | 117 | 78 |
| 2 | 86,195 | 86,755 | 561 | 2 | 34,470 | 58,893 | 24,424 | 78 |
| 3 | 86,818 | 86,912 | 95 | 3 | 58,956 | 59,297 | 342 | 76 |
| 4 | 87,013 | 87,242 | 230 | 4 | 59,400 | 59,784 | 385 | 76 |
| 5 | — | — | — | 5 | 59,827 | 59,937 | 111 | — |
| 6 | 87,330 | 87,430 | 101 | 6 | 59,983 | 60,375 | 393 | 82 |

Figure 1 Dot plot analysis of introns of SIAP3X and SIAP3Y. Each intron sequence of SIAP3X (vertical) is compared to that of SIAP3Y (horizontal). The first 700 bp of SIAP3Y intron 2 is shown. Numbers above each axis indicate the ordinal number of each intron. Yellow squares indicate comparisons for which homologous regions were detected.
The dates when the LTR retrotransposon–like elements were inserted differed depending on the kind of element. The two copia retrotransposon–like elements in the 13d11E sequence were assumed to be inserted 1.2 MYA (SE, 0.48) and 1.6 MYA (SE, 0.50). The one and two gypsy retrotransposon–like elements in the 13d11E and 7a8D sequences were thought to be inserted 0.22 MYA (SE, 0.16), 0.62 MYA (SE, 0.11), and 0.65 MYA (SE 0.12), respectively. The three unknown retrotransposon–like elements in 13d11E were considered to be inserted 0.18 MYA (SE, 0.057), 0.30 MYA (SE, 0.079), and 0.74 MYA (SE 0.12), respectively. One and two pairs of nested LTR retrotransposon–like elements were detected in 13d11E and 7a8D, respectively (Figure 5). It was natural to assume that the inner element was inserted after the outer element. With regard to the two pairs in the 7a8D sequence, the estimated insertion dates of the inner elements were more recent than those of the outer elements. The estimated insertion time of the inner element of the pair identified in 13d11E was older than that of the outer element, but the two insertion dates were determined to be close and not significantly different.

**Homology between noncoding DNA regions of the X and Y chromosomes**

The noncoding DNA regions around SlAP3X and SlAP3Y—except for the transposons that were inserted after recombination stopped—should have been homologous. We attempted to reveal structural changes of the sex chromosomes after recombination suppression by investigating how much homologous sequence remained. To detect homologous regions between the X and Y chromosomes, we conducted BLASTN searches of 13d11E (SlAP3X) using 7a8D (SlAP3Y) as the query sequence. Homologous regions were scattered around SlAP3X/Y (Table 3 and Figure 6A). The total length of the homologous regions in the 13d11E and 7a8D sequences were 7.7 kb (6.7%) and 7.7 kb (10.6%), respectively. We calculated the nucleotide divergence of each homologous sequence pair (Table 3). The divergences showed some variation, ranging between 0.085 (SE, 0.0085) (Table 3, ID 4) and 0.25 (SE, 0.029) (Table 3, ID 5), whereas that observed for the coding regions of SlAP3X/Y was 0.12 (SE, 0.034).

To investigate whether the homologous regions between 13d11E and 7a8D in the noncoding DNA regions were unique in the genome, we performed genomic Southern hybridizations with probes made from each PCR product of the homologous region in the 13d11E sequence (Table 3 and Figure 6B). Although probe sequences do not include transposons, all the hybridizations produced smeary patterns, which implied that these sequences were abundant in the genome. FISH performed using one of those sequences (Figure 6B, I) as a probe also yielded many signals on multiple chromosomes (Figure 6C). To investigate the uniqueness of noncoding DNA regions in the genome, which did not show homology to their counterparts on the opposite chromosome (Figure 6A, horizontal lines), we also analyzed genomic Southern hybridization data using PCR products generated from these DNA regions as probes (Figure 6B, a–j). All hybridizations yielded smeared patterns, which indicated that the genome included a large number of these sequences. FISH performed using one of these sequences (Figure 6B, d) as a probe also showed signals on multiple chromosomes, which supported the repetitive nature of the sequence (Figure 6D).

**DISCUSSION**

**Beginning of divergence between sequences surrounding SlAP3X and SlAP3Y**

We estimated that recombination between SlAP3X and SlAP3Y stopped 3.4 MYA. In addition, we determined the dates of the subsequent

---

**Figure 2** Genomic Southern blot analysis of SlAP3X and SlAP3Y. (A) Schematics of genomic sequences of SlAP3X and SlAP3Y. White and black boxes indicate untranslated regions and exons, respectively. Numbers indicate intron numbers of the genes and bold lines indicate probe sequence coverage of each intron for Southern hybridization. (B) Genomic Southern blot analysis of introns of SlAP3X and SlAP3Y. Numbers above lanes indicate intron numbers of the genes corresponding to those in (A).

**Figure 3** Schematics of BAC sequences. Horizontal lines indicate entire sequences of each BAC clone drawn to scale as indicated. Black bars indicate exons, which are labeled with Roman numerals.
insertions of transposons in this region. The molecular clock used for age estimations, which was the same as the one Nicolas et al. (2005) used, was based on the Brassicaceae species, which is currently the most reliable clock for S. latifolia. More accurate estimation will be possible when a molecular clock of more closely related species becomes available.

To confirm the location of SlAP3X within the nonrecombining region, we constructed a linkage map of the S. latifolia X chromosome based on the estimation of genetic distances (Table S7). The most probable order of the four genes from the PAR on the X chromosome was SlX1, SlAP3X, DD44X, and SlX4. The relative locations of genes on the X chromosome, integrated with the data presented by Bergero et al. (2013), are shown in Figure S1. Twenty percent divergence is typical for stratum 1 genes and 10% divergence is typical for stratum 2 genes (Cegan et al. 2010). SlAP3X was mapped between DD44X and SlX1, and the divergence between SlAP3X and Y was 12%, which implies that SlAP3X is located in stratum 2.

Differentiation of sequences in sex chromosomes

Many diverged transposons were found in 13d11E (SlAP3X) and in 7a8D (SlAP3Y) (Figure 4); however, the sequence contents of these elements were determined to be different. For example, no significant similarity was found between retroelements identified in 13d11E (SlAP3X) and 7a8D (SlAP3Y), indicating either that these retroelements were independently inserted into 13d11E (SlAP3X) or 7a8D (SlAP3Y) after recombination stopped or that they had existed on the proto-sex chromosomes and were translocated to a distant region on only one of the chromosomes. Because the estimated insertion time of the oldest LTR retroelement (1.6 MYA) is more recent than the predicted time when recombination between SlAP3X and SlAP3Y (3.4 MYA) stopped, the former scenario is more likely to have occurred.

Gschwend et al. (2012) compared the sequence of the C. papaya X chromosome with the autosomal counterpart of the related species, Vasconcellea monoica, and reported that sequences not only on the Y chromosome but also on the X chromosome expanded after recombinbination suppression. Obtaining the sequence of the SlAP3 ortholog

| Clone | ID | Type | Start, bp | End, bp | Length, bp | Left LTR Length, bp | BLASTN | K (SE) | Estimated Insertion Time, MYA (SE) |
|-------|----|------|-----------|---------|------------|---------------------|--------|-------|-----------------------------------|
| 13d11E (SIAP3X) | 1 | Unknown | 8721 | 14,467 | 5956 | 1548 | 0 | 99 | 0.0065 (0.0021) | 0.18 (0.057) |
| 2 | Unknown | 23,393 | 29,638 | 6246 | 1387 | 0 | 99 | 0.011 (0.0028) | 0.30 (0.079) |
| 3 | Gypsy | 54,004 | 70,569 | 10,886 | 715 | 0 | 99 | 0.0080 (0.0057) | 0.22 (0.16) |
| 4 | Unknown | 58,455 | 64,134 | 5680 | 1450 | 0 | 96 | 0.027 (0.0044) | 0.74 (0.12) |
| 5 | Copia | 74,749 | 79,514 | 4766 | 169 | 7E-65 | 93 | 0.045 (0.017) | 1.2 (0.48) |
| 6 | Copia | 88,141 | 92,782 | 4642 | 208 | 5E-86 | 95 | 0.057 (0.018) | 1.6 (0.50) |
| 7 | Unknown | 12,651 | 32,108 | 14,093 | 366 | 4E-153 | 94 | 0.044 (0.012) | 1.2 (0.32) |
| 8 | Unknown | 21,888 | 27,252 | 5365 | 1670 | 0 | 99 | 0.0030 (0.0013) | 0.083 (0.037) |
| 9 | Gypsy | 36,730 | 58,367 | 12,663 | 1538 | 0 | 97 | 0.022 (0.0039) | 0.62 (0.11) |
| 10 | Gypsy | 45,305 | 54,279 | 7770 | 1345 | 0 | 97 | 0.024 (0.0043) | 0.65 (0.12) |

K, nucleotide divergence between left and right LTRs; BLASTN, homology between the left and right LTR sequences calculated by the BLASTN program.
and its flanking region in Silene vulgaris, which is the closest related hermaphroditic species to S. latifolia, would enable us to better discern changes in the DNA sequence that occurred independently in the X and Y chromosomes.

All three unknown retrotransposon-like elements (ID 1, ID 2, and ID 4 in Table 2 and Figure 4) in 13d11E (SIAP3X) showed evidence of homology. In addition, the sequences of the two DNA transposon-like elements, which were nested in 13d11E (SIAP3X), were also homologous to one another (Figure 4). Thus, each group of transposons was concluded to have the same origin. The estimated insertion times of the three unknown retrotransposon-like elements in 13d11E (SIAP3X) were similar (0.18–0.74 MYA) (Table 2), suggesting that transposons of this type were translocated within a limited period. The estimated insertion times of the two copia retrotransposon-like elements in 13d11E (SIAP3X) were also similar (1.2 and 1.6 MYA) (Table 2). It is possible that each group of transposons has a shared origin but has lost homology over time. LINEs were detected only in the 7a8D sequence (Figure 4); however, according to the mapping of transposons by FISH analyses (Kejnovsky et al. 2009), they were distributed across the full length of both the X and Y chromosomes. It is possible that LINEs could be identified in the adjacent sequence of 13d11E (SIAP3X).

Yu et al. (2008) compared two pairs of BAC sequences derived from sex chromosomes of C. papaya. One pair of sequences (X BAC 61H02 and Yh BAC 95B12) was thought to have stopped recombining 7.0 MYA, in that 81.6% and 74.5% of sequences of the X and Yh chromosome had homology, respectively, and the identity between homologous sequences was 87.4%. The other pair of sequences (X BAC 53E18 and Yh BAC 85B24) was estimated to have ceased recombination 1.9 MYA. In this example, 41.5% and 30.7% of sequences of the X and Yh chromosome had homology, respectively, and the identity between homologous sequences was 83.6%. In the pair of sequences (13d11E and 7a8D) analyzed here for S. latifolia, which were assumed to have stopped recombining 3.4 MYA, 6.7% and 10.6% of sequences of the X and Yh chromosomes had homology, respectively, and the average identity between homologous sequences was 76.9%. Differences in the percentages of the noncoding homologous region are thought to be, in part, attributable to differences in the lengths of the analyzed sequences. The lower percentages despite the identity may reflect a high rate of accumulation of transposable elements, which may have partially caused the heteromorphism of sex chromosomes in this species.

Regarding indels, Sundström et al. (2003) conducted comparative analysis on intron sequences of paralogs in primates and birds and revealed that in birds with female heterogamety, indels occurred in the Z chromosome approximately twice as often as in the W chromosome, and that in primates with male heterogamety, indels occurred at similar rates in both the X and Y chromosome. This indicated that both meiotic recombination and the number of cell divisions in reproductive cells play an important role in the occurrence of indels. It raises an interesting question regarding whether this theory is also applicable to plants in which reproductive cells are generated from

**Table 3** Homologous regions between peripheral regions of SIAP3X and SIAP3Y

| ID | 13d11E (SIAP3X) | 7a8D (SIAP3Y) | Direction | E-value | Identity, % | K (SE) |
|----|----------------|---------------|-----------|---------|------------|-------|
| 1  | 3586 | 5587 | 2002 | 71,985 | 70,005 | 1981 | Inverted | 0 | 73 | 0.14 (0.012) |
| 2  | 16,253 | 17,744 | 1492 | 8055 | 9530 | 1474 | Direct | 0 | 76 | 0.11 (0.010) |
| 3  | 17,735 | 18,992 | 1258 | 3248 | 1996 | 1253 | Inverted | 0 | 81 | 0.14 (0.014) |
| 4  | 31,730 | 33,237 | 1508 | 1996 | 3467 | 1470 | Direct | 0 | 78 | 0.085 (0.0085) |
| 5  | 96,491 | 97,215 | 725 | 67,473 | 66,715 | 759 | Inverted | 5.00E–166 | 78 | 0.25 (0.029) |
| 6  | 96,492 | 97,219 | 728 | 13,016 | 13,789 | 772 | Direct | 4.00E–161 | 77 | 0.092 (0.013) |

K, nucleotide divergence between the regions in 13d11E and 7a8D.
vegetative cells, thus weakening the influence of the difference in the number of cell divisions between male and female reproductive cells. However, regarding transposable elements, a comparison between the 40-kb neo-X and 45-kb neo-Y sequences of \textit{Drosophila miranda}, which is assumed to have emerged approximately 1 MYA, revealed that insertions of seven transposable elements occurred only in the neo-Y chromosome (Bachtrog 2005). It is assumed that transposable elements tend to accumulate more in the recombination suppression region of the Y chromosome. In the regions in \textit{S. latifolia} analyzed in this study, transposable elements have highly accumulated not only in the Y chromosome but also in the X chromosome for 3.4 million years, preventing us from revealing whether the insertion of the transposable element was Y chromosome–biased. Obtaining continuous sequences between \textit{SlAP3} and its neighboring genes in both X and Y chromosomes will enable us to find the answers to our questions.

Origin of the noncoding DNA region

All sequences of noncoding DNA regions investigated in this study contained repetitive elements, irrespective of whether they were homologous between sex chromosomes (Figure 2B and Figure 6, B–D). Five noncoding DNA regions exhibited homology between the X and Y chromosomes (Table 3, IDs 1–4 and ID 6) and showed no significant differences in divergence from the synonymous nucleotide divergences observed for the translated regions of \textit{SlAP3X} and \textit{SlAP3Y}, as determined by the chi-square test. Because the divergence estimates for these regions along with \textit{SlAP3X/Y} are consistent with their chromosomal location in stratum 2, it is considered that these regions are derived from ancestral chromosomes. However, one noncoding DNA region (Table 3, ID 5) had higher divergences ($P=0.0001$ by chi-square test). Although the cause of the high divergence is uncertain, it is presumed that the sequences of the region have paralogous origins that had been duplicated in the past. Alternatively, as suggested in primates (Iwase et al. 2010), it could be possible that gene conversion occurred locally between the \textit{SlAP3X} and \textit{SlAP3Y} coding regions after recombination suppression, which would make divergence lower. With regard to noncoding DNA regions that were not homologous between the sex chromosomes, it was assumed that these sequences were inherited from ancestral chromosomes and that their counterparts in either the X chromosome or the Y chromosome were lost during deletion events.

Evolutionary strata on the X chromosome in \textit{S. latifolia}

Bergero et al. (2013) revealed that the X chromosome of \textit{S. latifolia} have three evolutionary strata. However, because whole sequences of sex chromosomes have not been generated, it is uncertain whether the
evolutionary strata on the X chromosome of *S. latifolia* have clear boundaries. Lemaître et al. (2009) proposed that evolutionary strata on the human X chromosome were formed by inversions on the Y chromosome, which has been explained by the isochromatid model (Ranz et al. 2007), because boundary segments between strata 4 and strata 5 and between PAR and strata 5 on the X chromosome are duplicated on the Y chromosome. According to the isochromatid model, two pairs of staggered single-strand breaks yield long 5’ overhangs, which are filled-in by DNA synthesis. When such breaks are repaired by nonhomologous end-joining, they can result in inversions flanked by inverted duplications of the sequences between the paired single-strand breaks (Ranz et al. 2007).

In two instances, we identified noncoding DNA regions that were present at two locations on one sex chromosome and only one homologous location on the second sex chromosome (Figure 6A). For example, region 3/4 in 7a8D (SlAP3Y) was homologous to region 3 and region 4 in 13d11E (SlAP3X). Region 2 in 7a8D, which was adjacent to region 3/4, was also homologous to region 2 in 13d11E, which was adjacent to region 3 (Figure 6A). These homologous noncoding DNA regions could be vestiges of inversions; however, definitive boundaries of sequence similarity that could be used to indicate evolutionary strata were not found.

In this study, we analyzed 116 kb and 73 kb of sequence from the X and Y chromosomes of *S. latifolia* and found that only 6.7% of the X chromosome–derived sequence and 10.6% of the Y chromosome–derived sequence were homologous between the two chromosomes. It was shown that sex chromosomes of *S. latifolia* were heteromorphic not only at the chromosomal level but also at the sequence level. We detected two locations in which homologous sequences between the X and Y chromosomes were duplicated, implying the occurrence of inversions. A wider comparison of noncoding DNA regions between the X and Y chromosomes, especially those near the PAR, will help delineate the boundaries of evolutionary strata and define differences in the degree of heteromorphy between these strata, which will be essential for gaining a broader understanding of the evolution of plant sex chromosomes.

**ACKNOWLEDGMENTS**

We thank Yoko Kuroki and Asao Fujiyama for technical support as a part of the support research project conducted by MEXT KAKENHI “Genome Science” (grant 221S0002). This work was supported by JSPS KAKENHI (grant 09073777 to K.I., grants 20780009, 23770070, and 25292009 to Y.K., and grants 18567001 and 24657046 to S.K.), by incentive research grants from RIKEN (to K.I.), by the Fund for Seeds of Collaborative Research from RIKEN (to Y.K.).

**LITERATURE CITED**

Asakawa, S., I. Abe, Y. Kudoh, N. Kishi, Y. Wang et al., 1997 Human BAC library: construction and rapid screening. Gene 191: 69–79.

Bachtrog, D., 2005 Sex chromosome evolution: Molecular aspects of Y-chromosome degeneration in *Drosophila*. Genome Res. 15: 1393–1401.

Bergero, R., A. Forrest, E. Kamau, and D. Charlesworth, 2007 Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. Genetics 175: 1945–1954.

Bergero, R., S. Qiu, A. Forrest, H. Borthwick, and D. Charlesworth, 2013 Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. Genetics 194: 673–686.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos et al., 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Cegan, R., G. A. B. Marais, H. Kubekova, N. Blavet, A. Widmer et al., 2010 Structure and evolution of *Apetala3*, a sex-linked gene in *Silene latifolia*. BMC Plant Biol. 10: 180.

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

Charlesworth, D., 2013 Plant sex chromosome evolution. J. Exp. Bot. 64: 405–420.

Charlesworth, D., and B. Charlesworth, 1980 Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. Genet. Res. 35: 205–214.

Du, J., Z. Tian, C. S. Hans, H. M. Luten, S. B. Cannon et al., 2010 Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. Plant J. 63: 584–598.

Edgar, R. C., 2004 MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32: 1792–1797.

Gschwend, A. R., Q. Yu, E. J. Tong, F. Zeng, J. Han et al., 2012 Rapid divergence and expansion of the X chromosome in papaya. Proc. Natl. Acad. Sci. USA 109: 13716–13721.

Ishii, K., R. Sugiyama, M. Onuki, Y. Kazama, S. Matsunaga et al., 2008 The Y chromosome-specific STS marker MS2 and its peripheral regions on the Y chromosome of the dioecious plant *Silene latifolia*. Genome 51: 251–260.

Ishii, K., Y. Amanai, Y. Kazama, M. Ikeda, H. Kamada et al., 2010 Analysis of BAC clones containing homologous sequences on the end of the Xq arm and on chromosome 7 in the dioecious plant *Silene latifolia*. Genome 53: 311–320.

Iwase, M., Y. Satta, H. Hirai, H. Hirai, H. Imai et al., 2003 The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. Proc. Natl. Acad. Sci. USA 100: 5258–5263.

Iwase, M., Y. Satta, H. Hirai, H. Hirai, and N. Takahata, 2010 Frequent gene conversion events between the X and Y homologous chromosomal regions in primates. BMC Evol. Biol. 10: 225.

Jurka, J., V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany et al., 2000 Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110: 462–467.

Kazama, Y., R. Sugiyama, S. Matsunaga, F. Shibata, W. Uchida et al., 2003 Organization of the *Kpsl* family of chromosomal distal-end satellite DNAs in *Silene latifolia*. J. Plant Res. 116: 317–326.

Kazama, Y., R. Sugiyama, Y. Suto, W. Uchida, and S. Kawanoe, 2006 The clustering of four subfamilies of satellite DNA at individual chromosome ends in *Silene latifolia*. Genome 49: 520–530.

Kejnovsky, E., R. Hobza, T. Cermak, Z. Kubat, and B. Vyskot, 2009 The role of repetitive DNA in structure and evolution of sex chromosomes in plants. Heredity 102: 533–541.

Koch, M., B. Haubold, and T. Mitchell-Olds, 2000 Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis, Arabis* and related genera (Brassicaceae). Mol. Biol. Evol. 17: 1483–1498.

Kosambi, D. D., 1944 The estimation of map distance from recombination values. Ann. Eugen. 12: 172–175.

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

Lawson-Handley, L. J., H. Cepilis, and H. Ellegren, 2004 Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics 167: 367–376.

Lemaître, C., M. D. Braga, C. Gautier, M. F. Sagot, E. Tannier et al., 2009 Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. Genome Biol. Evol. 1: 56–66.

Liu, Z., P. H. Moore, H. Ma, C. M. Ackerman, M. Ragiba et al., 2004 A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427: 348–352.

Matsunaga, S., M. Hizume, S. Kawanoe, and T. Kuroiwa, 1994 Cytological analyses in *Melandrium album*: genome size, chromosome size and fluorescence in situ hybridization. Cytologia (Tokyo) 59: 135–141.

Matsunaga, S., S. Kawanoe, H. Takano, H. Uchida, A. Sakai et al., 1996 Isolation and developmental expression of male reproductive organ-specific genes in a dioecious campion, *Melandrium album* (*Silene latifolia*). Plant J. 10: 679–689.
Matsunaga, S., S. Kawano, and T. Kuroiwa, 1997 MROS1, a male stamen-specific gene in the dioecious campion Silene latifolia is expressed in mature pollen. Plant Cell Physiol. 38: 499–502.
Matsunaga, S., E. Isono, E. Kejnovsky, B. Vyskot, J. Dolezel et al., 2003 Duplicative transfer of a MADS box gene to a plant Y chromosome. Mol. Biol. Evol. 20: 1062–1069.
Ming, R., S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte et al., 2008 The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature 452: 991–996.
Mosolek, M., P. Pasierbek, J. Malarz, M. Mos, and A. J. Joachimiak, 2005 Rumex acetosa Y chromosomes: constitutive or facultative heterochromatin? Folia Histochem. Cytobiol. 43: 161–167.
Nam, K., and H. Ellegren, 2008 The chicken (Gallus gallus) Z chromosome contains at least three non-linear evolutionary strata. Genetics 180: 1131–1136.
Nei, M., and S. Kumar, 2000 Molecular Evolution and Phylogenetics, Oxford University Press, Oxford, England.
Nicolas, M., G. Marais, V. Hykelova, B. Janousek, V. Laporte et al., 2005 A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biol. 3: e4.
Nishiyama, R., K. Ishii, E. Kifune, Y. Kazama, K. Nishihara et al., 2010 Sex chromosome evolution revealed by physical mapping of SIAP3X/Y in the dioecious plant Silene latifolia. Cytologia (Tokyo) 75: 319–325.
Ranz, J. M., D. Maurin, Y. S. Chan, M. von Grothhuss, L. W. Hillier et al., 2007 Principles of genome evolution in the Drosophila melanogaster species group. PLoS Biol. 5: e152.
Renner, S. S., and R. E. Ricklefs, 1995 Dioecy and its correlates in the flowering plants. Am. J. Bot. 82: 596–606.
Rice, P., I. Longden, and A. Bleasby, 2000 EMBOSS: The European Molecular Biology Open Software Suite. Trends Genet. 16: 276–277.
Široký, J., M. A. Lysák, J. Doležel, E. Kejnovský, B. Vyskot et al., 2001 Heterogeneity of rDNA distribution and genome size in Silene spp. Chromosome Res. 9: 387–393.
Sundström, H., M. T. Matthew, and H. Ellegren, 2003 Insertion and deletion mutation male biased?: Molecular evolutionary analysis of avian and primate sex chromosome sequences. Genetics 164: 259–268.
Tamura, K., and M. Nei, 1993 Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10: 512–526.
Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei et al., 2011 MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.
Telgmann-Rauber, A., A. Jamari, M. S. Kinney, J. C. Pries, and C. Jung, 2007 Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Mol. Genet. Genomics 278: 221–234.
Wang, I., J. K. Na, Q. Yu, A. R. Gschwend, J. Han et al., 2012 Sequencing papaya X and Y chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl. Acad. Sci. USA 109: 13710–13715.
Yamato, K. T., K. Ishizaki, M. Fujisawa, S. Okada, S. Nakayama et al., 2007 Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. USA 104: 6472–6477.
Yu, Q., S. Hou, F. A. Feltus, M. R. Jones, J. E. Murray et al., 2008 Low X/Y divergence in four pairs of papaya sex-linked genes. Plant J. 53: 124–132.

Communicating editor: N. D. Young