Substitution Rates in the X- and Y-Linked Genes of the Plants, 
Silene latifolia and S. dioica

Dmitry A. Filatov* and Deborah Charlesworth†

*School of Biosciences, University of Birmingham; and †Institute of Cell, Animal, and Population Biology, 
University of Edinburgh

Theory predicts that selection should be less effective in the nonrecombining genes of Y-chromosomes, relative to 
the situation for genes on the other chromosomes, and this should lead to the accumulation of deleterious nonsyn-
ynomous substitutions. In addition, synonymous substitution rates may differ between X- and Y-linked genes 
because of the male-driven evolution effect and also because of actual differences in per-replication mutation rates 
between the sex chromosomes. Here, we report the first study of synonymous and nonsynonymous substitution rates 
on plant sex chromosomes. We sequenced two pairs of sex-linked genes, SIX1-SIX1 and SIX4-SIX4, from dioecious 
Silene latifolia and S. dioica, and their non-sex-linked homologues from nondioecious S. vulgaris and Lychnis flos-
jovis, respectively. The rate of nonsynonymous substitutions in the SIX4 gene is significantly higher than that in the 
SlX4 gene. Silent substitution rates are also significantly higher in both Y-linked genes, compared with their X-
linked homologues. The higher nonsynonymous substitution rate in the SIX4 gene is therefore likely to be caused 
by a mutation rate difference between the sex chromosomes. The difference in silent substitution rates between the 
SIX4 and SlY4 genes is too great to be explained solely by a higher per-generation mutation rate in males than 
females. It is thus probably caused by a difference in per-replication mutation rates between the sex chromosomes. 
This suggests that the local mutation rate can change in a relatively short evolutionary time.

Introduction

Sex chromosomes are well known as an extreme instance of differences in gene density and recombin-
ation rates. The Y chromosomes of many species are ge-
ettically degenerate. Animal Y chromosomes carry 
many fewer genes than the ancestrally homologous X 
chromosomes (Lucchesi 1978; Lahn and Page 1999), 
though in plants with Y chromosomes it is not yet 
known whether these have lost many genes that are pre-

sent on the X. Y chromosome degeneration is probably 
an evolutionary consequence of the lack of recombin-
ation, which makes natural selection less effective than 
in recombining regions of genomes (Charlesworth 1994; 
Charlesworth and Charlesworth 2000). When recombi-
nation is rare or absent, nonindependence of selection 
at different loci reduces the effective population size and 
may allow slightly deleterious substitutions that would 
otherwise be eliminated (Charlesworth 1994; Charles-
worth and Charlesworth 2000; McVean and Charles-
worth 2000). Deleterious mutations will accumulate and 
become fixed by the stochastic process of Muller’s ratchet (stochastic loss of chromosomes with the fewest 
mutations; Gordo and Charlesworth 2000) and also by hitchhiking processes associated with the spread of advan-
tageous mutations (selective sweeps; Rice 1987) and 
elimination of deleterious ones (background selection; 
Charlesworth 1994). In contrast, X chromosomes are 
gene rich and recombine in female meiosis, and natural 
selection may be even more effective than on the au-
tosomes because of exposure of recessive mutations in 
hemizygous males (Charlesworth, Coyne, and Barton 
1987).

The processes just described for Y chromosome de-

generation all reduce effective population sizes for Y-
linked genes, compared with the X chromosome. This 
should lead to reduced sequence diversity in degener-
ating Y chromosomes and to reduced efficacy of selec-
tion in Y-linked coding sequences (Charlesworth and 
Charlesworth 2000). The efficacy of selection can be 
examined by studying silent and nonsynonymous site 
divergence between homologous genes of related spe-
cies. Assuming neutrality of mutations at silent sites, the 
substitution rate estimates the mutation rate (Kimura 
and Ohta 1971). Such sites should diverge in proportion to 
the product of the mutation rate and divergence time, 
and differences in silent site divergence between genes 
will reflect differences in their mutation rates (Kimura 
1983).

If selection is less effective for Y chromosomal 
genes, nonsynonymous site divergence (Ka) should be 
higher, relative to divergence at silent sites (Ks), in 
the coding sequences of Y-linked than X-linked genes. We 
should thus find higher Ka/Ks values for Y-linked genes. 
It is difficult to test for faster accumulation of deleteri-
ous mutations in Y-linked genes, because many animal 
Y chromosomes are almost completely degenerate and 
contain very few genes. Of the small number of known 
human genes and pseudogenes with both X- and Y- 
linked copies (Lahn and Page 1999), few homologous 
sequences from non-human species have been studied. 
Comparisons of sequences of the SMEX-SMCY pair of 
genes from humans, the horse, and the mouse yielded a 
Ka/Ks ratio for SMCY about an order of magnitude high-
er than that for SMEX, supporting a reduced efficacy of 
selection for SMCY (Agulnik et al. 1997). Much smaller 
differences were, however, reported for the avian sex-
linked CHD1 genes (Fridolfsson and Ellegren 2000). The 
human Y-linked DAZ gene, which has no X-linked 
homologue, can be compared with a homologous auto-
somal gene, DAZL1. The higher Ka/Ks ratio for DAZ 

Key words: sex chromosomes, efficacy of selection, mutation 
rates.

Address for correspondence and reprints: Dmitry A. Filatov, 
School of Biosciences, University of Birmingham, Edgbaston, Bir-

mingham B15 2TT, United Kingdom. E-mail: d.filatov@bham.ac.uk.

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038
could indicate reduced selective constraint (Agulnik et al. 1998), but a recent analysis suggests that the high $K_s/K_a$ ratio in the DAZ gene is caused by a few sites undergoing directional selection, whereas other sites are under purifying selection (Bielawski and Yang 2001).

High divergence between animal X and Y chromosomes (Lahn and Page 1999) makes $K_s$ and $K_a$ estimates unreliable. Interpretation of such results is also difficult because an observed high $K_s/K_a$ value for Y-linked genes may be caused by divergence in function (for example murine Zfy1 and Zfy2; see Johnston, Shimeld, and Sharpe 1998) rather than by lower efficacy of selection on the Y chromosome. Thus, for a reliable comparison of the substitution rates on the X and Y chromosomes, one should compare several homologous X- and Y-linked genes that are not too divergent from each other. Comparisons of five pairs of X- and Y-linked genes on the Drosophila miranda neo-Y and neo-X chromosomes have demonstrated higher $K_s/K_a$ on the neo-Y chromosome (Yi and Charlesworth 2000; Bachtrog and Charlesworth, 2002). Two of the Y-linked genes may have lost function because the $K_s/K_a$ ratio estimated from divergence from D. pseudoobscura is close to one, and no expression data are available (Yi and Charlesworth 2000). Three genes, however, have low $K_s/K_a$, suggesting that both neo-Y and neo-X copies are functional, yet $K_s/K_a$ on the neo-Y branch of the phylogeny is significantly higher than that on the branch to the neo-X (Bachtrog and Charlesworth, 2002).

The $K_s/K_a$ tests use relative nonsynonymous and silent site divergence values to compensate for possible differences in mutation rates between the two sex chromosomes, and it is also of interest to test for such differences. On the autosomes, deleterious recessive mutations are sheltered by a second copy of the gene in females, whereas in males the X is hemizygous, and such mutations are exposed to selection. It should therefore be advantageous to reduce the X chromosome mutation rate (Kondrashov 1995; McVean and Hurst 1997). Theoretical arguments indeed suggest that the selective advantage of modifiers reducing the mutation rate on the X chromosome is higher than that for the autosomes (McVean and Hurst 1997). On the other hand, Y chromosomes often carry few genes, and their mutation rate can probably increase without negative consequences for fitness. Synonymous substitution rates may therefore differ between the sex chromosomes. We have previously shown that the low diversity of the first Silene latifolia Y-linked gene studied (Sly1) is not attributable to a reduced Y chromosome mutation rate; rather, it appears to evolve faster than its X-linked homologue (Fi-latov et al. 2001), in line with this theory. This finding, however, implies that one cannot test for differences in the efficacy of selection without also testing for mutation rate differences.

In species with male heterogamety, Y chromosomes are confined to males; therefore, they experience more cell divisions per generation than X chromosomes because of a greater number of cell divisions in the male germ line (there are often more divisions per sperm than per egg). Although plants have no germ line (Poethig, Coe, and Johri 1986), more cell divisions occur in pollen than ovule production (Klekowski 1988). This may lead to the male-driven evolution effect: a higher mutation frequency in male than female gametes (Miyata et al. 1987; Shimmin, Chang, and Li 1993; Ellegren and Fridolfsson 1997; Smith and Hurst 1999). This alone can cause Y chromosomal genes to have higher mutation rates than X-linked ones. It is difficult to distinguish between this effect and a true difference in mutation rates (per replication) between the sex chromosomes because both processes lead to slower divergence of X-linked sequences than Y-linked ones (Smith and Hurst 1999). In birds, however, female heterogamety means that the two factors have opposing effects, and observations of a higher Z than W (female specific) chromosome mutation rate thus support a male-driven evolution effect (Ellegren and Fridolfsson 1997). In species with male heterogamety, the two hypotheses can be discriminated by comparing silent substitution rates of Y- and X-linked genes ($R_Y$, $R_X$) and autosomal genes ($R_A$), because autosomes are less restricted to females than is the X. It has been shown that under male-driven evolution, $R_Y/R_X \leq 3$. and $R_Y/R_A \geq 2/3$ (Miyata et al. 1987). On the basis of comparisons of X-linked and autosomal sequences from different species, a $R_Y/R_A$ value of 0.62 was estimated for rodents. The observed reduction in the X-chromosomal silent substitution rate for rodents is therefore too great to be explained by male-driven evolution, suggesting a true lower per-replication mutation rate of X-linked genes (McVean and Hurst 1997; Smith and Hurst 1999). Y-linked genes did not differ significantly from the autosomes, though there are few data (McVean and Hurst 1997). A recent analysis of divergence between mouse, rat, and human orthologues found chromosome-specific mutation rates, with the X-chromosome mutation rate lower than that of most, though not all, of the autosomes (Lercher, Williams, and Hurst 2001).

Here, we describe several tests for reduced efficacy of selection in Y-linked genes of the dioecious plant species S. latifolia and S. dioica, including analyses of codon usage in homologous X- and Y-linked genes. We also compare silent substitution rates on the sex chromosomes and find a higher rate on the Y than the X chromosome. Our results suggest that male-driven evolution is not the cause of the observed substitution rate differences. Per-replication mutation rates may therefore differ between the sex chromosomes, as in rodents.

The two species used in this study have chromosomal sex determination (XY males and XX females; Westergaard 1959). The species are close relatives (Goulson and Jerrim, 1997) and have morphologically similar sex chromosomes (Westergaard 1959). Silent site divergence between the two species is low for the two pairs of sex-linked genes for which extensive sequences are so far available (1.4% and 2.7% for the SIX1 and Sly1 genes, respectively, and 1.3% and 1.2% for SIX4 and Sly4), whereas divergence of these genes in either species from the homologues in nondioecious relatives is much higher (table 1 and fig. 1). The S. latifolia and S. dioica sex chromosomes probably evolved before the
Table 1
Estimates of Pairwise Divergence (D) Between the Two S. latifolia and S. dioica Sex-Linked Genes and the Non-Sex-Linked Homologues from S. vulgaris. The Divergence Estimates are Jukes-Cantor Distances ± Standard Error (SE) per 100 Sites. N stands for the Number of Sites in Each Pairwise Comparison, and Subscripts of the D and N Parameters Indicate Types of Sites (t stands for total sites, i for intron sites, and sites in coding regions are denoted by a and s, for amino acid replacement and synonymous sites, respectively).

<table>
<thead>
<tr>
<th>Species, Gene</th>
<th>Nt</th>
<th>Dt ± SE</th>
<th>Ni</th>
<th>Di ± SE</th>
<th>Ns</th>
<th>Ds ± SE</th>
<th>Ns</th>
<th>Ds ± SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>S. latifolia, SIX1 . . .</td>
<td>3167</td>
<td>6.5 ± 0.4</td>
<td>2759</td>
<td>7.8 ± 0.6</td>
<td>668.44</td>
<td>1.2 ± 0.4</td>
<td>189.56</td>
<td>7.0 ± 2.0</td>
</tr>
<tr>
<td>S. latifolia, SIY1 . . .</td>
<td>3635</td>
<td>7.1 ± 0.5</td>
<td>2777</td>
<td>8.5 ± 0.6</td>
<td>668.44</td>
<td>1.2 ± 0.4</td>
<td>189.56</td>
<td>8.7 ± 2.2</td>
</tr>
<tr>
<td>S. dioica, SIX1 . . .</td>
<td>3707</td>
<td>6.5 ± 0.4</td>
<td>2749</td>
<td>7.7 ± 0.5</td>
<td>668.44</td>
<td>1.2 ± 0.4</td>
<td>189.56</td>
<td>7.0 ± 2.0</td>
</tr>
<tr>
<td>S. dioica, SIY1 . . .</td>
<td>3622</td>
<td>7.1 ± 0.5</td>
<td>2764</td>
<td>8.6 ± 0.6</td>
<td>668.44</td>
<td>1.2 ± 0.4</td>
<td>189.56</td>
<td>8.1 ± 2.1</td>
</tr>
<tr>
<td>S. latifolia, SIY4 . . .</td>
<td>1734</td>
<td>5.8 ± 0.6</td>
<td>615</td>
<td>10.6 ± 1.4</td>
<td>856.89</td>
<td>1.4 ± 0.4</td>
<td>262.11</td>
<td>10.6 ± 2.0</td>
</tr>
<tr>
<td>S. latifolia, SIY4 . . .</td>
<td>1781</td>
<td>11.3 ± 0.9</td>
<td>665</td>
<td>18.7 ± 1.8</td>
<td>855.78</td>
<td>2.6 ± 0.6</td>
<td>260.22</td>
<td>23.6 ± 3.3</td>
</tr>
<tr>
<td>S. dioica, SIY4 . . .</td>
<td>1724</td>
<td>5.4 ± 0.6</td>
<td>605</td>
<td>10.1 ± 1.4</td>
<td>857.11</td>
<td>1.0 ± 0.3</td>
<td>261.89</td>
<td>10.8 ± 2.0</td>
</tr>
<tr>
<td>S. dioica, SIY4 . . .</td>
<td>1778</td>
<td>11.2 ± 0.8</td>
<td>665</td>
<td>18.5 ± 1.8</td>
<td>854.00</td>
<td>2.4 ± 0.5</td>
<td>259.00</td>
<td>24.3 ± 3.4</td>
</tr>
</tbody>
</table>

species separated. Consistent with this, silent site divergence between SIY4 and SIY4 is 15.5% for both dioecious species, much higher than that between the species. Between SIX1 and SIY1, however, divergence is only 2%, though SIX1 and SIY1 do not recombine (Filatov et al. 2000). Recombination between these genes may therefore have ceased recently (Atanassov et al. 2001). The difference in divergence between the SIX4-SIY4 and SIX1-SIY1 pairs of genes resembles the evolutionary strata discovered in human sex chromosomes (Lahn and Page 1999). One advantage of these species is the relatively low age of the sex chromosomes (Desfeux et al. 1996), so that divergence between homologous X- and Y-linked genes is low compared with animal sex chromosomes. The low divergence of the homologous X- and Y-linked genes, and the low Ks/Kt values in their coding sequences, allows us to assume that these genes have similar functions and experience similar selective pressures. Furthermore, most other species in the genus Silene do not have separate males and females (dioecy) and thus have no sex chromosomes (Desfeux et al. 1996). Thus, we can compare genes on the sex chromosomes with non-sex chromosomal homologues in nondioecious relatives, avoiding the difficulties of comparing different sets of genes between the sex chromosomes and autosomes (Smith and Hurst 1999).

Materials and Methods
Isolation of Genes

Genomic DNA was isolated from leaves of individual plants of each of the species as described earlier (Filatov and Charlesworth 1999). All PCR amplifications were done with Expand® High Fidelity PCR System (Boehringer Mannheim) and the following primer pairs (see fig. 1): XY1+1 and X1−7 for the S. latifolia SIX1 gene, XY1+1 and Y1−8 for the S. latifolia SIY1 gene, XY1+3 and X1−7 for the S. dioica SIX1 gene, Y1+18 and XY1−10 for the S. dioica SIY1 gene, XY1+1 and XY1−10 for the S. vulgaris and Lychnis flos-jovis gene homologous to SIX1 and SIY1, XY4+1

![Exon-intron structure of the genes studied showing the regions sequenced. The exon-intron structure was inferred from comparison of the genomic DNA and cDNA sequences (Delichere et al. 1999; Filatov et al. 2000; Atanassov et al. 2001). Regions sequenced in the present study are indicated, and positions and directions of the PCR and sequencing primers are shown by small arrows. The following PCR primers were used: XY1+1 GATGGGCGCCCTGTGGAGG, Y1−8 ACCCAAGTATCTTCTTCGCTCC, XY1+3 AGGCTCGTTCTCCCTTTGTG, X1−7 ACTGGCAACGACTTCATTTTGGG, Y1+18 CCTCTTTACCGACTTCATCTCCATGC, XY1−10 TCGAGGAGGGGGGACGGTC, XY4+1 ACAATTGATGAGTGAATAGT, X4−3 AAATTACGAAAGGCAGTAAAACGC, Y4−7 AATCACACAGTGGATGTCATTTTG, and XY4−14 GTGGTCAATCACACAGTGTATAC. Sequencing primers are available on request.](image-url)
Substitution Rates on Silene Sex Chromosomes

Fig. 2.—Neighbor-joining tree for fourfold degenerate and noncoding (mostly intron) sites of the (A) SIX1-SIX1 and (B) SIX4-SIX4 genes (2,618 and 699 sites, respectively) and their non-sex-linked homologues from S. vulgaris and L. flos-jovis. The branch lengths (Jukes-Cantor distances) were calculated using the baseml program (Yang 2001), assuming a separate substitution rate for each branch. Different shadings show locations of the genes: white for non-sex-linked genes, black for Y-linked, grey for X-linked genes. The branches correspond to the assignment of the three substitution rates in the baseml analysis and to the three $K/J$ ratios in the codeml analysis (see Materials and Methods). Branch lengths are omitted for very short branches.

and X4–3 for the S. latifolia and S. dioica SIX4 gene, XY4+1 and Y4–7 for the S. latifolia and S. dioica SIX4 gene, XY4+1 and XY4–14 for the homologue of SIX4 and SIXY in S. vulgaris and L. flos-jovis. SIX1, SIXY, SIX4, and SIXY are single copy genes in S. latifolia (Dellichere et al. 1999; Atanassov et al. 2001), and only one band was amplified from S. vulgaris and L. flos-jovis for these genes. The amplification products were cloned using the TOPO TA Cloning Kit (Invitrogen) and sequenced using an ABI Prism 377 automatic sequencer (Perkin-Elmer) and 29 specific primers (available on request). The ABI chromatogram sequence files were corrected, and contigs were assembled using ProSeq v2.9 (Filatov 2001). The sequences were aligned by the ClustalW Version 1.8 program (Thompson, Higgins, and Gibson 1994). Alignment of a 0.15-kb AT-rich region from intron 12 of the SIX1-SIX1 was uncertain because of multiple insertions and deletions, and this region was removed from the final alignment (fig. 1).

GeneBank accession numbers for the sequences reported in this paper: AY084036, AY084037, AY084038, AY084039, AY084040, AY084041, AY084042, AY084043, AY084044, AY084045, AY084046, AY084047.

Nucleotide Substitution Rates

Pairwise divergence values and the topology of the gene trees were calculated using MEGA software (Kumar, Tamura, and Nei 1994). These topologies were used in the maximum likelihood (ML) analyses described subsequently. In all the ML analyses assuming a local molecular clock (Yoder and Yang 2000), substitution rates were assigned to the different branches shown in figure 2. Likelihood ratio (LR) tests (Muse and Weir 1992) were used to test the significance of differences in substitution rates. LR tests, and the ML estimates of the relative substitution rates, were done with the baseml program from the PAML Version 3e package (Yang 2001). To compare nonsynonymous substitution rates in the SIX4 and SIXY genes, we used the codeml program from the PAML Version 3e package (Yang 2001) to extract all first and second codon positions from the data set for these genes, for use in the LR test. Similarly, to compare silent substitution rates in the X- and Y-linked genes, we performed LR tests on noncoding and fourfold degenerate sites. To obtain the ML estimates of the ratios of the substitution rates $(R_U/R_X, R_U/R_A)$ and their standard errors, we assigned a sequence whose rate is the denominator to be the ancestral sequence (rate “#0” in baseml notation; see Yang 2001) and used the local molecular clock (Yoder and Yang 2000) with three substitution rates assigned as shown in figure 2. The relative rates estimated by the baseml program estimate the $R_U/R_X$ or $R_U/R_A$ ratios depending on whether the X-linked or the autosomal sequence was assigned as ancestral.

To estimate $K/J$ ratios for the X- and the Y-linked genes, the codeml program from the PAML Version 3e package (Yang 2001) was used, assuming the phylogeny in figure 2. Each branch was assumed to have a separate substitution rate (no clock mode), and three $K/J$ ratios were assigned as shown in figure 2 by different branch shadings. For the LR test, the model with three $K/J$ ratios was compared with one with just two ratios, one for autosomal and one for the X- and the Y-linked genes.

Amino Acid Substitution Patterns

To compare the amino acid substitution patterns in the SIX4 and SIXY genes, we used the approach of Wyckoff, Wang, and Wu (2000), using the codon mutation patterns tool of the ProSeq Version 2.9 software (Filatov 2001). As the S. dioica and S. latifolia amino acid sequences are nearly identical, only the S. latifolia-coding sequences were analyzed. Assuming the tree topology shown in figure 2B, the sequence of the gene ancestral to SIX4 and SIXY was reconstructed using parsimony (Fitch 1971). Synonymous substitutions in the SIX4 and SIXY genes were counted, and nonsynonymous substitutions were classified into four classes according to the Grantham (1974) distance ($D$): conservative ($D < 51$), moderate ($50 < D < 101$), radical ($100 < D < 151$), and very radical ($D > 150$). The expected numbers ($\pm$Standard error [SE]) of mutations in each class were obtained in 5,000 repetitions of the following simulation (Wyckoff, Wang, and Wu 2000). Mutations were introduced as follows: the site of each mutation was chosen randomly and its nucleotide mutated to any other nucleotide, assuming transitions to be 1.5 times more frequent than transversions (based on the average transition-transversion ratio for the SIX4 and SIXY genes). The ancestral coding sequence was mutated until the number of mutations at either synonymous sites (re-
ferred to subsequently as “expected S”) or nonsynonymous sites (“expected N”) reached the observed value.

Codon Bias

To detect preferred codons in *Silene*, we followed the approach proposed by Akashi (1995). We used 36 *S. latifolia* cDNA sequences from GenBank that have open reading frames (8,353 codons in total; accession numbers: M16887, X94358, Y12529, Y12603, X02432, Y08773, Y08774, Y08775, Y08776, Y08777, Y08778, Y08779, Y08780, Y12324, Y12325, AB000108, D82026, D82027, AB013611, AB013612, AF062480, Y08155, X02965, M16888, X80488, X80489, X80490, X80491, X80492, Y18517, Y18519, U53828, U53829, U53827, AJ310658, AJ310657). For these sequences, we calculated two measures of codon bias, the scaled U53827, AJ310658, AJ310657). For these sequences, we calculated two measures of codon bias, the scaled

Results

Replacement Substitutions

Figure 1 shows the regions of the two genes sequenced, and Figure 2 shows trees for the two genes studied. Both outgroups are consistent in the position of the node which roots the homologous X- and Y-linked sequences. Figure 2 also shows the estimated divergence values between the sequences. Between the \emph{SiX1} and \emph{SiY1} genes of either *S. latifolia* or *S. dioica*, there were no replacement site differences, but between \emph{SiX4} and \emph{SiY4}, \( K_s = 2.4 \pm 0.5\% \) for *S. latifolia* and \( 1.6 \pm 0.4\% \) for *S. dioica*. To compare the replacement substitution rates in these genes, we estimated their nonsynonymous site divergence from the homologous outgroup sequence of *S. vulgaris* (table 1). The nonsynonymous substitution rates are higher in the \emph{SiY4} gene than in \emph{SiX4}, for both *S. latifolia* and *S. dioica*, but the \emph{SiY4} synonymous rates are also higher (table 1). The \( K_s \) values for the \emph{SiX4} and \emph{SiY4} genes differ significantly. This conclusion was confirmed by ML analysis of the substitution rates in the first and the second coding positions \( (740 \) bp) of the \emph{SiX4}-\emph{SiY4} genes. A model with three rates (X-linked, Y-linked, and non–sex-linked; \( 2\Delta L = 12.52, P < 0.001 \) by \( \chi^2 \) with \( df = 1 \)) fits the data significantly better than the model with two rates (with X- and Y-linked genes having the same rate and different non–sex linked rate).

As explained previously, however, different nonsynonymous substitution rates in the \emph{SiX4} and \emph{SiY4} genes may result from either differences in the efficacy of selection in these genes or different mutation rates. The higher silent substitution rate in \emph{SiY4} than \emph{SiX4} suggests that the high \emph{SiY4} nonsynonymous rate may not indicate reduced efficacy of selection. To compensate for different mutation rates, we therefore compared \( K_s/K_a \) ratios for the X- and Y-linked genes. The ratio for \emph{SiY4} \( (K_s/K_a = 0.119 \pm 0.0019) \) was only slightly higher than that for the \emph{SiX4} gene \( (K_s/K_a = 0.115 \pm 0.0040) \), and the difference is not significant. LR test (see Materials and Methods), was also nonsignificant \( (2\Delta L = 0.0) \).

Evolution of the X- and Y-Linked Amino Acid Sequences

If the efficacy of selection is lower in Y- than X-linked genes, another prediction is that Y-linked genes should accumulate more substitutions of nonconservative amino acids. We therefore compared the types of amino acid replacements in the \emph{SiY4} and \emph{SiX4} genes by estimating the numbers and types of mutations in the coding region of the *S. latifolia* \emph{SiX4} and \emph{SiY4} genes since they became separated (table 2). *Silene dioica* sequences are very similar to those of *S. latifolia*, and nearly all nonsynonymous differences between \emph{SiX4} and \emph{SiY4} in *S. dioica* are identical to those in *S. latifolia*. This analysis was not conducted for the \emph{SiX1}-\emph{SiY1} genes because there are no nonsynonymous substitutions between these genes.

Fourteen amino acid replacements were detected in the *S. latifolia* \emph{SiY4} gene but only three in \emph{SiX4} (table 2), which limits the sensitivity of the analysis. For both \emph{SiX4} and \emph{SiY4}, the number of replacements in all non-
Table 2
Substitutions in the Coding Regions of the SIX4-SIY4 Genes. Data are Partitioned Into Synonymous and Four Nonsynonymous Substitution Classes According to the Physicochemical Differences Represented by Grantham Distances (D) (Grantham 1974) Between the Amino Acids. Expected Numbers of Substitutions (± standard deviation) Were Calculated, Given the Observed Number of Synonymous (S) or Pooled Nonsynonymous (N) Sites

<table>
<thead>
<tr>
<th>Nonsynonymous</th>
<th>Observed SlY4 S. latifolia</th>
<th>Expected (S)</th>
<th>Expected (N)</th>
<th>Observed SlX4 S. latifolia</th>
<th>Expected (S)</th>
<th>Expected (N)</th>
<th>Observed SlY4 S. vulgaris</th>
<th>Expected (S)</th>
<th>Expected (N)</th>
<th>Observed SlX4 S. vulgaris</th>
<th>Expected (S)</th>
<th>Expected (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synonymous</td>
<td>36</td>
<td>36 ± 0</td>
<td>6.11 ± 2.976</td>
<td>7</td>
<td>7 ± 0</td>
<td>13 ± 0</td>
<td>2.18 ± 1.750</td>
<td>4</td>
<td>13 ± 0</td>
<td>2.18 ± 1.750</td>
<td>4</td>
<td>13 ± 0</td>
</tr>
<tr>
<td>(D &lt; 51)</td>
<td>(50 &lt; D &lt; 101)</td>
<td>(100 &lt; D &lt; 151)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
<td>(D &gt; 150)</td>
</tr>
<tr>
<td>Moderate</td>
<td>4</td>
<td>4.19 ± 1.743</td>
<td>0.89 ± 0.794</td>
<td>1</td>
<td>1.034</td>
<td>1.31 ± 1.374</td>
<td>0.86 ± 0.794</td>
<td>1.415</td>
<td>1.31 ± 1.374</td>
<td>0.86 ± 0.794</td>
<td>1.415</td>
<td>1.31 ± 1.374</td>
</tr>
<tr>
<td>Radical</td>
<td>6</td>
<td>6.05 ± 1.880</td>
<td>1.29 ± 0.857</td>
<td>3</td>
<td>2.42 ± 1.415</td>
<td>2.19 ± 1.111</td>
<td>0.87 ± 0.847</td>
<td>1.34 ± 1.113</td>
<td>2.19 ± 1.111</td>
<td>0.87 ± 0.847</td>
<td>1.34 ± 1.113</td>
<td>2.19 ± 1.111</td>
</tr>
<tr>
<td>Very Radical</td>
<td>3</td>
<td>14.32 ± 4.517</td>
<td>2.81 ± 2.029</td>
<td>1</td>
<td>2.42 ± 1.415</td>
<td>5.28 ± 2.766</td>
<td>0.29 ± 0.513</td>
<td>1.55 ± 1.381</td>
<td>5.28 ± 2.766</td>
<td>0.29 ± 0.513</td>
<td>1.55 ± 1.381</td>
<td>5.28 ± 2.766</td>
</tr>
<tr>
<td>Conservative</td>
<td>1</td>
<td>8.11 ± 3.307</td>
<td>1.55 ± 1.381</td>
<td>1</td>
<td>1.55 ± 1.381</td>
<td>2.95 ± 1.931</td>
<td>0.29 ± 0.513</td>
<td>0.29 ± 0.513</td>
<td>2.95 ± 1.931</td>
<td>0.29 ± 0.513</td>
<td>0.29 ± 0.513</td>
<td>2.95 ± 1.931</td>
</tr>
</tbody>
</table>

Synonymous categories (see Materials and Methods) was lower than that expected from simulations conditioned on the observed number of silent substitutions in these genes (Wyckoff, Wang, and Wu 2000), in agreement with their low Ks/Ka ratios. The observed distribution of amino acid replacements also did not differ significantly from that expected, conditioning the simulations on the observed numbers of nonsynonymous substitutions (pooling the four nonsynonymous classes), suggesting that selection has not altered the pattern of replacements in the SIY4 gene.

Codon Bias

Low efficacy of selection for Y-linked genes also predicts reduced codon bias in Y- compared with X-linked genes. We cannot compare segregating and fixed synonymous substitutions on the X and Y chromosomes (Akashi 1995) because there are few sequence variants on the X-linked and Y-linked genes. We cannot compare segregating and fixed synonymous substitutions on the X and Y chromosomes (Wyckoff, Wang, and Wu 2000), in agreement with their low Ks/Ka ratios. The observed distribution of amino acid replacements also did not differ significantly from that expected, conditioning the simulations on the observed numbers of nonsynonymous substitutions (pooling the four nonsynonymous classes), suggesting that selection has not altered the pattern of replacements in the SIY4 gene.

SIX4. The excess of unpreferred over preferred substitutions in the SIY4 gene is marginally significant (G-test = 3.85, P < 0.05). Using all 18 codons positively correlated with codon bias, regardless of significance, we find 12 unpreferred and seven preferred synonymous substitutions in SIY4 (G-test = 1.33, NS) versus one preferred in the SIX4 gene. There are only six synonymous differences between SIY1 and SIX1, so this analysis is not possible for these genes.

Silent Substitution Rates

As noted previously, the intron and synonymous site divergence of the SIY4 gene from its homologue in the nondioecious S. vulgaris are significantly higher than that in SIX4 (table 1). A similar, but less pronounced, effect is found for SIY1 and SIX1 (table 1). To test the significance of the substitution rate differences in the branches of the gene trees for X- and Y-linked and autosomal loci (fig. 1), we did LR tests (Muse and Weir 1992) using the local molecular clock mode of the base.ml program (Yang 2001). We estimated the likelihood of the null model with three evolutionary rate parameters, one for each of the X-linked, Y-linked, and non-sex-linked branches (fig. 2). This was compared with the likelihoods of the following alternative models with two rate parameters (table 3): (1) X- and Y-linked genes forced to have the same rate, (2) X-linked and non-sex-linked genes rates forced to be equal, and (3) Y-linked and non-sex-linked genes having the same rate. The three-parameter model fits the data significantly better than any of the two-parameter models (table 3). Thus, silent substitution rates differ significantly between the X- and Y-linked and non-sex-linked branches (fig. 2). This was compared with the likelihoods of the following alternative models with two rate parameters (table 3): (1) X- and Y-linked genes forced to have the same rate, (2) X-linked and non-sex-linked genes rates forced to be equal, and (3) Y-linked and non-sex-linked genes having the same rate. The three-parameter model fits the data significantly better than any of the two-parameter models (table 3). Thus, silent substitution rates differ significantly between the X- and Y-linked and non-sex-linked branches (fig. 2). This was compared with the likelihoods of the following alternative models with two rate parameters (table 3): (1) X- and Y-linked genes forced to have the same rate, (2) X-linked and non-sex-linked genes rates forced to be equal, and (3) Y-linked and non-sex-linked genes having the same rate. Therefore, the three-parameter model is preferred.

Male-Driven Evolution

To test for male-driven evolution, we estimated RY/RX and RY/RX ratios and their SE by ML, assuming the gene tree topologies in figure 2 and using the model with
three substitution rates described previously. All the ratios were estimated for silent sites, using only fourfold degenerate and noncoding (mostly intron) sites (2,618 sites for the SlY1-SlX1 genes, and 699 sites for the SlY4-SlX4 genes). R_A/R_X is 0.25 ± 0.06 for SlY1-SlX1 and 0.47 ± 0.15 for SlY4-SlX4, both significantly below 2/3. For SlY4-SlX4, R_A/R_X is 4.24 ± 1.18, marginally significantly greater than three. Thus, the difference in substitution rates for these genes is difficult to explain by male-driven evolution. For the SlY1 and SlX1 genes, however, R_A/R_X is 1.91 ± 0.39, well within the range predicted under male-driven evolution.

Possible Causes of Mutation Rate Differences

In mammals, the substitution rates of different genomic regions have been reported to correlate with their GC contents (Wolfe, Sharp, and Li 1989), though this was recently questioned (Hurst and Williams 2000). The possibility of different GC contents of the X- and Y-linked sequences is therefore worth examining. However, GC content is similar in all the species compared (data not shown). We also found no significant nonstationarity of the GC content of the genes studied, using the approach proposed by Eyre-Walker (1994).

Methylation is another possible cause of substitution rate differences between X- and Y-linked and autosomal genes. The observed numbers of CG pairs were significantly lower than that expected (in brackets) from the overall GC content. For S. latifolia, the values were, for SlY4 and SlX4, respectively: 42 (78), 47 (78), and for S. dioica SlY4 and SlX4 42 (76) and 49 (78). The deficiency of CG pairs is even greater in the SlX4 and SlY1 genes: 52 (123), 56 (124), 60 (125), 58 (124) for S. latifolia SlY1 and SlX1, and S. dioica SlY1 and SlX1 genes, respectively. This suggests that methylation inflates the number of C→T transitions in these genes. If the higher mutation rate in Y-linked genes is caused by overmethylation of the Y chromosome, C→T mutations should be a higher proportion of all mutations in CG dinucleotides, in Y-linked compared with X-linked genes. To test this, we used the reconstructed ancestral sequences (see Materials and Methods) to estimate these proportions for the X- and Y-linked genes. For the two dioecious species, the average proportion of C→T transitions in CG dinucleotides is higher in the X-linked genes (8.3% for SlX1 and 12.5% for SlX4) than in their Y-linked homologues (8.1% for SlY1 and 8.9% for SlY4). Thus, although methylation is an important mutagenic factor in these genes, the difference in substitution rates between the Y and the X chromosomal genes is not attributable to methylation.

To compare mutation patterns in the Y- and X-linked genes, we also counted the numbers of all types of mutations by comparing the ancestral and the actual sequences. The difference in mutation pattern between the homologous X- and Y-linked genes was significant neither for all sites nor for just silent sites. However, the transition (ts)/transversion (tv) ratio for SlY4 was much lower (1.24 ± 0.005) than that for the SlX4 gene (1.82 ± 0.016). A similar trend was observed in the SlX1 (ts/tv = 2.14 ± 0.015) and the SlY1 (ts/tv = 1.88 ± 0.009) pair of genes. The reasons for these differences in the substitution pattern are unclear, but this is the opposite of what would be expected if the difference were caused by C→T transitions in CG dinucleotides caused by methylation differences. The number of (C or G)→(A or T) substitutions is not significantly different from the number in the opposite direction (2 × 2 contingency G-test), corroborating our conclusion that GC content in these genes does not deviate significantly from stationarity.

Discussion

Selection on Genes in the X and Y Chromosomes

As explained previously, the effective population size, N_e, in the nonrecombining regions of Y chromosomes is predicted to be reduced, compared with that in the homologous X-linked genes. The consequent lower effectiveness of both positive and negative selection is thought to be the main cause of the genetic degeneration of the Y chromosome (Charlesworth and Charlesworth 2000). Previously, we demonstrated greatly reduced DNA diversity in the SlY1 gene compared with the SlX1 (Filatov et al. 2000, 2001). DNA diversity is about an order of magnitude lower in the Y-linked, compared with the orthologous X-linked gene, in S. latifolia and S. dioica. Similar results have been obtained for the SlY4 and SlX4 genes (V. Laporte, D. Filatov, and D. Charlesworth, unpublished data). A substantially reduced efficacy of natural selection is therefore also expected in Y-linked genes.

Whether reduced effective population size leads to a higher nonsynonymous replacement rate depends on the distribution of the selection coefficients on such mu-

Table 3

| Log-Likelihood Ratios for the Comparison of the Null Model, with Three Substitution Rates, and the Three Models with Two Rates (see text). The Significance is Tested Using the Approximation That the Log-Likelihood Ratios are Chi-square-Distributed (Muse and Weir 1992). The Chi-square Values Have One Degree of Freedom, as the Models Differ by One Parameter |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                 | **ALL SITES**   | **INTRONS AND four-fold DEGENERATE SITES** | **INTRON SITES** |
|                 | 1               | 2               | 3               | 1               | 2               | 3               |
| SlY1-SlX1 ...  | 11.5****        | 15.4***         | 7.4**           | 11.6***         | 16.1***         | 7.7**           |
| SlX4-SlY4 ...  | 62***           | 7.8**           | 21.5***         | 42.9***         | 5.5*            | 13.6***         |
|                 |                 |                 |                 | 10.46***        | 15.8***         | 7.9**           |

* P < 0.025; ** P < 0.01; *** P < 0.001.
tations, which is unknown (Keightley 1998). If selection against amino acid replacements is strong enough, the 10-fold reduction in the effective population size on the Y chromosome could still leave $N_e$ sufficiently large that selection can maintain genes’ amino acid sequences, accounting for our finding of no significant difference in the distribution of amino acid replacements between the $\text{SIX4}$ and $\text{SIY4}$ genes.

To detect a reduced effectiveness of natural selection, it may thus be most helpful to analyze weakly selected substitutions. In both $\text{SIX4}$ and $\text{SIY4}$, even conservative amino acid replacements accumulate less than expected under neutrality (table 2). Selection acting on synonymous codons should, however, be weak. For $D. \text{melanogaster}$ the selection coefficient was estimated to be about $10^{-5}$ (Akashi 1995). We indeed detect a significant excess of unpreferred over preferred synonymous substitutions in $\text{SIY4}$. However, this is only marginally significant, and there are too few substitutions in $\text{SIX4}$ to compare the X- and Y-linked homologues. More data are needed, both to establish the preferred codons in $\text{Silene}$ and to have sufficient data from sex-linked genes to draw more reliable conclusions. It seems unlikely, however, that a reduced Y-chromosome effective population size can explain both the higher synonymous and nonsynonymous divergence for Y-linked genes. The relationship between $N_e$ and expected divergence is nonlinear (see Akashi, Kliman, and Eyre-Walker 1998). It is therefore unlikely that nonsynonymous and synonymous substitution would be affected to the same degree, so this hypothesis cannot account for the absence of a higher $K_s/K_e$ ratio in $\text{SIY4}$ compared with the $\text{SIX4}$ gene.

Our findings therefore suggest that a higher mutation rate on the $\text{Silene}$ Y chromosome may be responsible for the higher number of amino acid replacements in $\text{SIY4}$, rather than reduced efficacy of selection. In principle, it is possible to distinguish these possibilities using diversity and divergence data (McDonald and Kreitman 1991; Akashi 1995). Unfortunately, the diversity on the Y chromosome is too low (Filatov et al. 2000, 2001), and this analysis will not be possible without large amounts of sequence data to provide sufficient variants.

**Mutation Rates on the X and Y Chromosomes**

It is clear that the silent substitution rate in the Y-linked genes is higher than that in the X-linked homologues (figure 2), and LR tests demonstrate the significance of the differences in silent substitution rates between the X-linked, Y-linked, and non-sex-linked genes (table 3). The most parsimonious explanation for the differences in silent substitution rates is that the mutation rates differ. Assuming that the ancestral mutation rate for the X- and Y-linked genes was the same as the rate in the non-sex-linked genes in $S. \text{vulgaris}$ and $L. \text{flos-jovis}$, our results show a reduction in both X-linked genes and acceleration in $\text{SIY4}$. Even without this assumption, the X-linked rate is significantly lower than in the Y-linked genes, and at least for the $\text{SIY4-SIX4}$ gene, this is difficult to explain solely by male-driven evolution.

If the effect observed in $\text{Silene}$ is caused by a higher number of cell divisions during pollen compared with ovule production, it must affect both $\text{SIX1-SIY1}$ and $\text{SIX4-SIY4}$ gene pairs. The difference between $\text{SIX1}$ and $\text{SIY1}$ is much lower than that between the $\text{SIX4}$ and $\text{SIY4}$ genes, however, suggesting that male-driven evolution is not the chief cause of the difference in silent substitution rates. Furthermore, for the substitution rate on the Y chromosome to be three times higher than that on the X chromosome, an enormous excess of cell divisions (and mutations) in males, compared with females, is required (Miyata et al. 1987). This might occur in wind-pollinated plants, which produce huge amounts of pollen (though this is not largely caused by extra cell divisions per generation). It is certainly not plausible for $S. \text{latifolia}$ and $S. \text{dioica}$, which are pollinated mostly by moths (Goulson and Jerrim 1997) and have moderate pollen production. Lower mutation rates on the X chromosome, and higher rates on the Y chromosome, have often been interpreted as resulting from male-driven evolution (Shimmin, Chang, and Li 1993; Huang et al. 1997). Our results suggest that this effect is unlikely to be the main cause of the different silent substitution rates between the $\text{Silene}$ sex chromosomes. Although we cannot rule out a contribution from male-driven evolution, it seems likely that the differences between the X- and Y-linked genes’ substitution rates reflect different per-replication mutation rates.

The reasons for different mutation patterns and rates between different regions of genomes are poorly understood (Matassi, Sharp, and Gautier 1999; Eyre-Walker and Hurst 2000; Lercher, Williams, and Hurst 2001), and the same is true for $\text{Silene}$. We cannot account for the observed substitution rate difference between the X- and the Y-linked genes in terms of different GC content (Wolfe, Sharp, and Li 1989; Hurst and Williams 2000). Although recombination itself may be mutagenic (Strathern, Shafer, and McGill 1995) and there are, of course, differences in the recombinational environment between the X- and the Y-linked genes, the trend is in the opposite direction from that required to explain the mutation rate difference: the nonrecombinating Y chromosome has the higher mutation rate. The $S. \text{latifolia}$ Y chromosome is more methylated than the X chromosome (Vyskot et al. 1993), which could lead to more C→T transitions caused by deamination of 5-methyl cytosine (Robertson and Wolfe 2000). Our analysis demonstrated that this hypothesis also does not explain the difference in mutation rates between the X- and the Y-linked genes. Interestingly, the fairly recent origin of the $\text{Silene}$ sex chromosomes (Desfeux et al. 1996) and the moderate divergence between the homologous X- and Y-linked genes in the dioecious species, imply mutation rate changes within a few million years.

**Acknowledgments**

D.C. was supported by the Natural Environment Research Council of Great Britain, D.A.F. by a grant to
D.C. from the Leverhulme Trust and by start-up funds from the University of Birmingham.

LITERATURE CITED


Smith, N. G. C., and L. D. Hurst. 1999. The causes of synonymous rate variation in the rodent genome: can substitution rates be used to estimate the sex bias in mutation rate? Genetics 152:661–673.


