Testing for the Footprint of Sexually Antagonistic Polymorphisms in the Pseudoautosomal Region of a Plant Sex Chromosome Pair

Suo Qiu,[2] Roberta Bergero,[2] and Deborah Charlesworth[1,2]
Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom

ABSTRACT The existence of sexually antagonistic (SA) polymorphism is widely considered the most likely explanation for the evolution of suppressed recombination of sex chromosome pairs. This explanation is largely untested empirically, and no such polymorphisms have been identified, other than in fish, where no evidence directly implicates these genes in events causing loss of recombination. We tested for the presence of loci with SA polymorphism in the plant *Silene latifolia*, which is dioecious (with separate male and female individuals) and has a pair of highly heteromorphic sex chromosomes, with XY males. Suppressed recombination between much of the Y and X sex chromosomes evolved in several steps, and the results in Bergero *et al.* (2013) show that it is still ongoing in the recombining or pseudoautosomal, regions (PARs) of these chromosomes. We used molecular evolutionary approaches to test for the footprints of SA polymorphisms, based on sequence diversity levels in *S. latifolia* PAR genes identified by genetic mapping. Nucleotide diversity is high for at least four of six PAR genes identified, and our data suggest the existence of polymorphisms maintained by balancing selection in this genome region, since molecular evolutionary (HKA) tests exclude an elevated mutation rate, and other tests also suggest balancing selection. The presence of sexually antagonistic alleles at a locus or loci in the PAR is suggested by the very different X and Y chromosome allele frequencies for at least one PAR gene.

IN the evolution of sex chromosomes, a major process is recombination suppression between the evolving Y and X chromosomes in the region containing the sex-determining genes. In many organisms with sex chromosome systems, including mammals, birds, and plants, recombination suppression has occurred in multiple events, at different times during the chromosomes' evolutionary history. Each event changes a formerly recombining region into a new nonrecombining region, shrinking the recombining pseudoautosomal region (PAR). These events created a series of so-called "evolutionary strata" that are recognized by their differing levels of divergence between X- and Y-linked gene pairs of mammals (Lahn and Page 1999; Skaletsky *et al.* 2003; Lemaitre *et al.* 2009), and of ZW genes in birds (Lahn and Page 1999; Skaletsky *et al.* 2003; Lawson-Handley *et al.* 2004; Matsubara *et al.* 2006; Nishida-Umehara *et al.* 2007; Nam and Ellegren 2008).

The evolutionary forces driving repeated changes in the boundary between the nonrecombining, male-specific region of the Y and the recombining PAR in different taxa are, however, not yet well understood. An attractive hypothesis (but not the only possibility, see *Discussion*) is that, as Y or W chromosomes evolve, sexually antagonistic alleles arise on these chromosomes, benefitting one sex, but harming the other. These alleles may become polymorphic in populations, with the Y (in XY systems) carrying male-benefit alleles at higher frequencies than the X, establishing a situation that selects for reduced recombination with the fully sex-linked sex-determining region (Charlesworth and Charlesworth 1980; Rice 1987; Mank and Ellegren 2009; Jordan and Charlesworth 2012). In a species with ongoing sex chromosome evolution, the PAR is the genome region most likely to maintain such "sexually antagonistic" (SA) polymorphisms, because partially sex-linked SA alleles can most readily maintain linkage disequilibrium (LD) with the linked

doi: 10.1534/genetics.113.152397
Manuscript received February 26, 2013; accepted for publication April 20, 2013
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152397/-/DC1.
Sequence data from this article have been deposited with NCBI under accession nos. KF019374-KF019629.
[1]Corresponding author: Ashworth Lab, King's Bldgs., West Mains Rd., University of Edinburgh, Edinburgh, EH9 3JT, United Kingdom. E-mail: deborah.charlesworth@ed.ac.uk
[2]All authors contributed equally to this work.

sex-determining region of the XY pair, producing sex differences in allele frequencies (Otto *et al.* 2011; Jordan and Charlesworth 2012).

A growing body of evidence suggests that loci with SA alleles exist in the genomes of various organisms (reviewed by Mank *et al.* 2008; Bonduriansky and Chenoweth 2009; Delcourt *et al.* 2009; Stulp *et al.* 2012). However, the only known SA polymorphisms are in fish, where male characters involved in attractiveness to females increase predation risk. These characters do not increase female fertility, so their expression in females is harmful (Lindholm and Breden 2002; Kitano *et al.* 2009; Roberts *et al.* 2009). Some of the genes controlling these characters are partially sex linked (Lindholm and Breden 2002), suggesting the potential for either selection favoring loss of recombination or to evolve sex-specific expression so that only males express the male-benefit traits, rendering tight Y-linkage unnecessary; some of these partially sex-linked genes indeed have male-specific expression. Polymorphisms at coloration loci in guppies tend to be restricted to sex-linked genes when predation is strong and the advantages and disadvantages in the two sexes conflict, but are more often found at autosomal loci when predation is weak (Gordon *et al.* 2012), suggesting that SA selection has affected the establishment of polymorphisms.

However, no evidence definitively connects SA alleles to the evolution of reduced recombination of Y or W sex chromosomes, and evidence is needed from systems other than fish sexual selection. Young sex chromosome systems, before multiple strata evolve, probably often have physically large PAR regions, containing many genes, and so such species should be ideal for testing these ideas. PAR genes of such species, including several plants (Liu *et al.* 2004; Mcdaniel *et al.* 2007; Spigler *et al.* 2008), could be used to test for signals indicating the presence of SA polymorphisms, particularly if recombination suppression is ongoing.

The plant *Silene latifolia* is suitable for studying the early stages of sex chromosome evolution. SA selection is plausible, because inflorescences and flowers show sexual dimorphism (*e.g.*, Delph *et al.* 2010; Delph and Herlihy 2012). Recombination suppression is estimated to have been initiated ~10 MYA (Bergero *et al.* 2007), *i.e.*, the system is much younger than those of mammals, and many genes retain functional Y-linked copies, *i.e.*, genetic degeneration of the *S. latifolia* Y chromosome is much less than in mammals or *Drosophila* (Bergero and Charlesworth 2011; Chibalina and Filatov 2011). Evolutionary strata are nevertheless present, indicating multiple recombination suppression events in *S. latifolia*'s ancestry (Bergero *et al.* 2007).

As a first step in testing the SA polymorphism hypothesis, we used a population genetic approach, because it does not require identifying the genes that are experiencing the hypothesized balancing selection, but relies on their "footprint" in the PAR, where genes exerting selection pressure for reduced recombination are expected to be found. If a PAR harbors SA polymorphism(s) in linkage disequilibrium with the sex-determining region of the XY pair, LD is expected with loci in the interval between the sexually antagonistic locus and the PAR boundary, and therefore genes in the intervening genome regions (and immediately distal PAR regions) should show elevated diversity.

In the accompanying article in this issue (Bergero *et al.* 2013), we mapped several genes to the *S. latifolia* PAR. The ~25-cM genetic map length of the *S. latifolia* PAR suggests that it probably carries many genes that could potentially undergo SA mutations, and so this species is well suited for testing the SA hypothesis, unlike species whose PARs include very few genes. Here, we analyze sequence diversity in a set of alleles from PAR genes. Our goals were twofold. The first is to further test for partial sex linkage, using population genetic data. The data were primarily collected for our second goal, which was to perform analyses to ask whether PAR genes show the expected footprints of polymorphism due to sexually antagonistic selection: high diversity, other evidence of balancing selection maintaining alleles polymorphic for a long evolutionary time, and evidence of LD between the PAR and the fully Y-linked, or male-specific, region (Patten *et al.* 2010; Úbeda *et al.* 2011).

Neutral sites very closely linked to any balanced polymorphism are expected to have higher diversity than surrounding genome regions (Hudson and Kaplan 1988). Even without SA selection, neutral diversity is thus expected to be elevated in the PAR (Kirkpatrick *et al.* 2010). However, recombination in each generation breaks down the associations causing elevated diversity, which become negligible unless the product of the recombination rate and the effective population size is small. This is consistent with the fact that the peaks of polymorphism near sites known to be under long-term balancing selection are often confined to the gene itself. For example, in the case of the primate ABO polymorphism, the diversity peak affects ~5 kb in and near exon 7 (Calafell *et al.* 2008; Ségurel *et al.* 2012). Therefore, under the neutral null hypothesis, only PAR genes very closely linked to the nonrecombining male-specific region of the Y chromosome (MSY) should be affected (Kirkpatrick *et al.* 2010). Even if the PAR contains a gene under SA selection, unless selection is extremely strong, X–Y allele frequency differences are not likely without close linkage to the MSY. In models of SA, only loci within <1 cM from the PAR boundary are expected to maintain balanced polymorphisms, and therefore only neutral variants within this distance, or slightly longer, should show LD with the MSY (Otto *et al.* 2011; Jordan and Charlesworth 2012). The findings in *Carica papaya* are consistent with these predictions. In the small fully sex-linked region (3.5 and 8.1 Mb in the X and Y, respectively), MSY genes are diverged from their X homologs, but sequence divergence decays within 80 kb of the border with the PAR (Wang *et al.* 2012); assuming a plausible recombination rate of 5 cM/Mb, this represents 0.4 cM. In *S. latifolia*, however, we find indications of associations with the MSY, including high diversity, for several PAR genes, all of which recombine with the fully sex-linked region, at least in some families, indicating that

the recombination rate for these genes in the population is at least several centimorgans (see Bergero *et al.* 2013). Below, we discuss possible explanations for our findings.

Materials and Methods

S. latifolia population samples for testing between partial and complete sex-linkage and estimating sequence diversity

To validate our genetic inferences of pseudoautosomal locations for some genes, and to study sequence diversity, we used a set of *S. latifolia* individuals of both sexes grown from seeds collected from naturally pollinated females from across Europe (Supporting Information, Table S1). Our sample of 12 males, one male per population, represent a subset of those sequenced in our previous comparison of fully X- and Y-linked genes (Qiu *et al.* 2010), with the addition of one male (from Ioanninon, Greece, sequenced only for the *E284* gene). The females were from the same populations (Table S1). As Table S1 shows, sequences could not be obtained for all plants for all the genes.

The *S. vulgaris* plants used for divergence estimates are the same as those used in our previous study of autosomal gene diversity in *S. latifolia* (Filatov *et al.* 2000, 2001; Matsunaga *et al.* 2003; Laporte *et al.* 2005; Qiu *et al.* 2010); plants from subspecies *vulgaris* (from Sussex, England, and from Scotland) and subspecies *maritima*, from the Channel Islands and an unknown location, were included.

DNA extraction, PCR reactions, and cloning

Genomic DNA for sequencing was extracted from leaves using the FastDNA kit (Qbiogene) following the manufacturer's instructions. The seedlings' sexes were determined by PCR amplification of intron 2 of the *SlCyp* gene, which contains a fixed 293-bp indel in the Y copy (Bergero *et al.* 2007).

The primers for PCR amplifications for five PAR loci and locus *E559*, which was an initial PAR candidate, are listed in Table S2. Alleles of these loci were sequenced from our natural population sample, together with three candidate PAR genes, *E559*, *E523*, and *E521*, which are apparently fully sex linked, based on the diversity survey (see Bergero *et al.* 2013). The PCR conditions in a Finnzymes' Piko cycler with Phire Hot-Start DNA Polymerase (Finnzymes) were as follows: 1 cycle of initial denaturation at 98° for 30 sec, 10 cycles of DNA denaturation at 98° for 5 sec, primer annealing varying from 60 to 70° for 5 sec, and DNA amplification at 72° for 30 sec, 25 cycles at 98° for 5 sec, 60° for 5 sec, 72° for 60 sec, and finally 1 cycle at 72° for 5 min. For genotyping polymorphic intron size variants, PCR products from parents and offspring were run on capillary electrophoresis according to Bergero *et al.* (2007).

To estimate sequence diversity of the PAR genes, the PCR products were cleaned with ExoSAP-IT (Amersham Biosciences, Tokyo) and sequenced on an ABI 3730 capillary sequencer (Applied Biosystems). Single nucleotide polymorphisms (SNPs) were obtained by direct sequencing from both strands. DNAs producing PCR amplicons with heterozygous indels in intron regions were reamplified with a proof-reading DNA polymerase (Phusion, Finnzymes) using the PCR conditions given above, and cloned into a T-tailed pBSKS+ vector (Stratagene) before sequencing. The final sequences were aligned in Sequencher 4.8 (Gene Codes, Ann Arbor, MI; http://www.genecodes.com), including sequences of the likely orthologous genes from *S. vulgaris*, and manually adjusted by using Se-al v. 2.0 (http://tree.bio.ed.ac.uk/software/seal/).

Sequence analyses

To study diversity, single-copy genes are essential, and this is particularly important because we find high diversity for several putative PAR genes (see *Results*). The genetic results in Bergero *et al.* (2013) show that most of the PAR genes studied are clearly single copy in the *S. latifolia* genome (section entitled "The *S. latifolia* X map and the PAR"). For two genes, *E241* and *E284*, however, some results suggested potential paralogues. The supporting text (File S1) describes these results, together with evidence that *E241* segregates as a single-copy gene. For *E284*, a paralogous copy was detected. One copy, denoted by $E284_{PAR}$, segregates as a PAR gene, and this was used in our diversity study, using locus-specific primers.

The aligned sequences of the putatively pseudoautosomal genes were first examined to see whether variants are shared between the Y and X chromosomes or are male-specific, indicating complete Y linkage. Male and female plants from the same natural populations (Table S1) were used to test for complete *vs.* partial sex linkage of all genes that are located within regions added to the sex chromosome pair since the formation of the younger of the two previously known evolutionary strata (Bergero *et al.* 2013).

Polymorphism analyses (including estimates of nucleotide diversity and Tajima's *D*) and divergence estimates between the *S. latifolia* and *S. vulgaris* sequences were done using DnaSP v.5.00.06 (Librador and Rozas 2009). The results for previously studied fully sex-linked genes in Table 1 are from published papers, only the last of which used the same natural population sample as for the PAR genes studied here (Filatov *et al.* 2000, 2001; Matsunaga *et al.* 2003; Laporte *et al.* 2005; Qiu *et al.* 2010). Y-linked alleles were distinguished from X-linked ones because they were found exclusively in males, enabling us to use Y- or X-specific PCR primers for sequencing.

We cannot estimate frequencies of alleles at most of the PAR loci studied, because the phase of variants is generally not known. Therefore we cannot directly test whether alleles differ in frequencies between the sexes. Instead, we used Tajima's *D*, which can detect an excess of intermediate frequencies expected under balancing selection (Tajima 1989). As will be shown below, this detects a strong signal for fully sex-linked loci, as expected since Y-specific variants will be at intermediate frequencies (1/4, assuming a 1:1 sex ratio). Although Tajima's *D* has not been explicitly modeled

Table 1 Shared and fixed X–Y SNP differences and nucleotide diversity, using all sites (because some loci include little coding sequence), but excluding indels

Gene	Numbers of sequences		Numbers of sites	Differences between the two sets		Silent site divergence of X from *S. vulgaris*	Nucleotide diversity (JC corrected)	
				Fixed differences	Shared variants		Males	Females
Fully sex-linked genes	Y	X						
SlXY4	45	40	888	67	0	0.153	0.124	0.004
SlXY9	23	23	243	3	0	0.149	0.121	0.0178
SlAp3XY	12	12	788	88	0	—	0.070	0.016
SlCypXY	48	46	1169	30	0	0.032	0.058	0.021
SlXY7	46	43	576	34	0	0.115	0.036	0.0053
DD44XY	14	13	989	25	0	0.113	0.031	0.015
SlXY1	22	24	1582	25	0	0.080	0.024	0.016
Totals	—	—	6235	272	0	—	—	—
Newly discovered fully sex-linked genes								
E559	13	27	595	13	6	0.077	0.007	0.002
E523	16	13	527	8	0	0.058	0.009	0.003
E521	4	7	640	21	0	0.075	0.019	0.003
PAR genes	♂	♀						
E219	24	16	426	0	48	0.161	0.046	0.028
E592	21	14	545	0	28	0.176	0.023	0.031
E200	23	18	297	0	11	0.054	0.004	0.004
E241	23	16	592	0	77	0.135	0.038	0.042
$E284_{PAR}$	33	15	588	0	109	0.093	0.068	0.067

For the fully sex-linked genes, all our sequences are from male plants, so we estimated diversity in females using just the X alleles; diversity in males is high for some of the fully sex-linked genes, due to the contribution from divergence of the Y sequences from those of the X-linked ones.

for partially sex-linked loci, it will clearly be elevated due to LD with the fully sex-linked region only in regions where diversity is elevated due to high coalescence times, *i.e.*, regions extremely closely linked to the MSY (see Introduction).

To estimate recombination rates within the PAR genes, we used LDhat software (McVean *et al.* 2002). To test LD directly, we first used HaploRec software (Eronen *et al.* 2006) to estimate phased haplotypes for the concatenated set of five PAR loci plus *E559*, using SNP variants with frequencies of at least 10%, since rare variants contribute little valuable information about LD. The resulting data were analyzed to test whether the number of haplotypes at a given locus is significantly lower than expected under neutrality. DnaSP was also used for these haplotype tests (DePaulis and Veuille 1998), using the numbers of variants and *R*/gene values estimated by the same program, and using DnaSP's coalescent simulation tool. We also used DnaSP to estimate the proportion of pairs of polymorphic sites with significant LD after Bonferroni correction, and the value of the LD measure Z_{ns} (Kelly 1997). To test the significance of diversity difference between sets of loci, excluding the effects of different mutation rates, HKA tests were done using a maximum likelihood implementation of the HKA test, MLHKA (Wright and Charlesworth 2004).

Results

Population genetic tests of PAR locations of candidate PAR genes

Before conducting population genetic tests of the SA polymorphism hypothesis, partially sex-linked genes require further testing to confirm that they are genuinely located in a PAR, because segregation data (as presented in Bergero *et al.* 2013) cannot always distinguish PAR genes and fully sex-linked genes that are closely linked to the PAR boundary; moreover genotyping errors (including bands from paralogous and other sequences) can potentially create the appearance of recombination when a gene is in fact completely sex linked. Population genetic evidence can, however, distinguish them. Fully sex linked genes (that never recombine between the X and Y chromosomes) will have variants fixed on the Y. Genes with multiple variants that are seen only in males can therefore reliably be assigned to the nonrecombining region of the XY pair. We therefore tested our putative PAR genes by sequencing them in a sample of male and female *S. latifolia* plants from natural populations (including the male parents of the three families in table 1 of Bergero *et al.* 2013).

Table 1 shows the distributions of fixed X–Y differences and shared variants in sequences obtained from the PAR genes sequenced and the fully sex-linked genes previously sequenced (see *Materials and Methods*) or newly sequenced in this study. All previously identified fully sex-linked genes have many fixed X–Y differences, and no shared variants, and measures of subdivision between X and Y sequences are high (Figure S1). In contrast, the six PAR genes show no male-specific variants or fixed X–Y differences, but many shared variants, consistent with partial sex-linkage and recombination in these genes, including *E352*, which appears fully sex linked in our mapping family, though not in two other families (Bergero *et al.* 2013).

This test also confirmed complete sex linkage for two other genes (that are closely linked to PAR genes: *E523* and

E521; see Bergero *et al.* 2013). Another such gene, *E559*, has numerous male-specific variants, supporting complete sex linkage; Y and X haplotypes were evident in each male plant of natural origin, while all females had X haplotypes (figure S2 of Bergero *et al.* 2013). However, we also found six polymorphisms shared between the Y and X haplotypes (Table 1). A PAR location therefore cannot be definitively excluded until detailed physical mapping becomes possible for this region of the *S. latifolia* genome. However, even with complete sex linkage, shared polymorphisms are not impossible. They would not be expected if recombination stopped due to an event in the genome region affected (*e.g.*, an inversion causing Y linkage of a PAR haplotype of the region), because diversity in the Y would be eliminated in the resulting selective sweep. However, the X- and Y-linked alleles could still recombine in regions away from inversion breakpoints, through double crossovers or by gene conversion, which has been documented in the male-specific regions of mammalian sex chromosome pairs (Slattery *et al.* 2000; Rosser *et al.* 2009; Trombetta *et al.* 2010; Ellegren 2011), particularly in the moderately recent stratum 4 (Iwase *et al.* 2010). However, the locations of the shared polymorphisms are not clustered, and do not suggest gene conversion. Alternatively, recombination suppression could occur with no inversion (e,g. through unlinked recombination modifiers Charlesworth *et al.* 1985; Ebinuma 1987; Perkins and Bojko 1992). In such situations, ancestral polymorphisms that were present before recombination stopped would be lost more slowly from the Y by genetic drift, and shared variants could persist for some time. The mechanism of recombination suppression between the *S. latifolia* Y and X is currently unknown.

The occurrence of homozygous males for a gene that shows evidence of sex linkage is a further indication that the gene is in a region that recombines with the sex-determining region. Genotype frequency results for our candidate PAR genes are consistent with this location. Homozygotes are not rare, and many are authenticated by genetic data in various families and cannot be explained as null alleles (Bergero *et al.* 2013).

Sex averaged recombination rate estimates for PAR genes are intermediate between those from fully X linked and autosomal genes (Figure S2). This is also consistent with partial sex linkage. Moreover, consistent with the data from families in Bergero *et al.* (2013), it does not suggest a very high recombination rate in males (which might be produced if the PAR is physically very small, and a crossover event occurs in every male meiosis).

Diversity in the PAR genes

Nucleotide diversity: To test for SA polymorphism in the PAR, we estimated sequence diversity for PAR genes and performed tests of the null hypothesis of neutrality.

Occasional failed PCR amplifications from some plants (Table S1) suggest that some alleles may have substitutions in the primer annealing sites. This problem was particularly severe for *E200* (indeed no alleles amplified from one individual, although other loci amplified from the same sample without problems). This gene is not devoid of variation; we found six length variants in intron 1 (which we did not attempt to sequence and align), and many indels in a different region that was sequenced. We omitted *E200* from our nucleotide diversity analyses, since nonamplified alleles probably cause underestimation of nucleotide diversity for this gene, and alignment problems are severe. We also did not sequence the *E352* gene, which also shows high allelic diversity, with multiple indel variants, making alignment impossible. Specifically, 9 different alleles, defined by their lengths, were detected in our total sample (or 11, including two further alleles in female plants of unknown origins). Four alleles were found in both sexes, 2 only in males or females (see Table S1; or 4, including the two females of unknown origins). The 254-bp variant in the Y chromosome of the parent of our mapping family (see Bergero *et al.* 2013) is common in males, but not male specific. This suggests that this gene has high diversity, but we cannot reliably estimate nucleotide diversity.

All four PAR loci for which we could obtain reliable estimates have very high nucleotide diversity (Table 1). For these loci, any potential underestimation of diversity would make this conclusion conservative. We analyzed all site types because, under the hypothesis being tested—that these genes might have high diversity due to LD with the sex-determining region on the Y chromosome—high diversity, reflecting LD, could be present throughout the sequence; the results are similar when diversity is estimated for silent sites (not shown). Despite these genes' fairly short sequences (Table 1), and the likelihood that alleles were missed in our set of sequences, as just explained, these genes also include many indel variants, which were often found in multiple individuals (eight in *E592*, three in *E219*, large numbers in $E284_{PAR}$ and *E241*, and six in *E200*).

A high mutation rate is one potential explanation for the high diversity of the subset of genes with PAR locations, and this possibility must be considered, as divergence from the orthologous *S. vulgaris* sequences is somewhat higher for PAR genes than for other genes (Table 1). HKA tests (which allow for the possibility of different mutation rates; see *Materials and Methods*) show that a high mutation rate is unlikely to explain the high diversity (Table 2). Compared with nine autosomal genes studied previously (see details in Qiu *et al.* 2010), the PAR genes have significantly ($P = 0.01$) higher within-species diversity than expected, except for *E200*, whose diversity is underestimated, as explained above (diversity in *E200* does not differ significantly from the four PAR genes whose diversity could be better estimated). However, the PAR genes' diversity does not differ significantly from that of the seven fully X-linked genes ($P = 0.06$), although the difference is marginally significant ($P = 0.043$) when we exclude the *E200* PAR gene.

LD and haplotype analyses: For fully sex-linked genes, diversity estimates are higher in males, as expected due to

Table 2 Nucleotide diversity estimates, diversity/divergence ratios, and results of MLHKA tests

	Divergence from *S. vulgaris*		Diversity/divergence from *S. vulgaris*	
Mean diversity ± SE	Silent sites	All sites	Silent sites	All sites
Autosomal genes	0.115 ± 0.010	0.039 ± 0.009	0.259 ± 0.076	0.365 ± 0.116
X-linked genes, excluding *E559*	0.107 ± 0.016	0.060 ± 0.001	0.192 ± 0.041	0.200 ± 0.043
X-linked genes, including *E559*(X)	0.103 ± 0.014	0.061 ± 0.009	0.191 ± 0.036	0.211 ± 0.037
PAR genes	0.141 ± 0.030	0.079 ± 0.009	0.447 ± 0.113	0.450 ± 0.113
HKA tests among PAR genes			*P* values	
	All 6 PAR genes		PAR genes excluding E559 (5 genes)	
E200	0.2		0.87	
E219	0.96		0.99	
E241	1		1	
E284	0.062		0.13	
E559	**0.011**		—	
E592	0.96		0.95	
Tests of the 5 PAR genes above (excluding *E559*) *vs.* other genes				
vs. 9 autosomal genes	—		**0.01**	
vs. 7 X-linked genes	—		0.06	
Tests of 4 PAR genes *vs.* other genes, excluding *E200* (see text)				
vs. 9 autosomal genes	—		**0.0063**	
vs. 7 X-linked genes	—		**0.043**	

Statistically significant results at the $P < 0.05$ level are in boldface type.

fixed differences between X and Y chromosomes (Table 1). Of the PAR genes, only *E352* shows a large sex difference in allele frequencies. Overall, 22/30 of the alleles in males had the common 254-bp variant, *vs.* 4/18 of the females' alleles; the sex difference in frequencies is significant by a 2 × 2 contingency table analysis ($P = 0.00019$ by a Fisher's exact test). For the other PAR genes, however, no such allele frequency differences were evident, and diversity is similar in both sexes. For *E241*, 12 different alleles could be distinguished by their lengths, and 2 were at high frequencies, but neither of these is restricted to one sex. Larger samples from within single populations, and analysis of sequences where phase is known, will be needed to better test allele frequency differences.

As a potentially more sensitive way to detect sex-specific variation in the four sequenced PAR genes, we also estimated net divergence between sequences from males and females (which will indicate associations of SNPs with the Y chromosome), and subdivision between males and females, using the population subdivision measure K_{ST}, which quantifies the proportion of variation that is between groups, rather than within groups, and whose statistical significance can be tested (Hudson *et al.* 1992). This approach gave no evidence for such subdivision for any of the PAR genes whose sequences could be analyzed (this excludes *E352* and *E200*, as explained above). *E559* and *E523*, which probably became fully sex-linked recently (see Bergero *et al.* 2013), yielded measures intermediate between the fully sex-linked genes and PAR genes (Figure S1). Overall, these tests give no evidence for alleles of these PAR loci being associated with the MSY.

Several other results reject the null hypothesis, but no PAR gene shows conclusive results for all tests. Haplotype tests (using the estimated haplotypes, see *Materials and Methods*) suggest that two PAR loci, *E241* and $E284_{PAR}$, have unexpectedly few haplotypes, consistent with balancing selection affecting these genes (Table 3). On the other hand, few pairs of sites yielded statistically significant LD (Table S3). Loci *E592* and *E219* gave significant results for the *H* test (conditioning the simulations on the value of θ per gene), again with unexpectedly few haplotypes, but *K* tests (conditioning on the value number of variants) were non-significant; nevertheless, a larger fraction of pairs of sites yielded significant tests for LD in these two genes. The alleles that could be analyzed from the *E200* gene gave no evidence for strong LD by any of these tests.

If balancing selection is affecting the PAR genes, positive Tajima's *D* values are expected (Tajima 1989). Given that autosomal genes or X-linked sequences alone yield generally negative values in *S. latifolia*, suggesting recent population expansion (Qiu *et al.* 2010), we tested whether PAR genes show values higher than other genes. The results are summarized in Figure 1. The means for autosomal and X-linked sequences are −0.53 ± 0.17 and −0.73 ± 0.16, respectively, based on all sites (results for silent sites alone are similar, and are not shown). In contrast, the fully sex-linked genes (pooling X and Y allele sequences from males) show the expected largely positive values (mean 1.04 ± 0.30), though the values are significant at $P < 0.05$ for only two loci; pooling all sequences, including X-linked alleles from females, gives similar results. The PAR genes (except for *E200*, which has few variants in our sequenced alleles, see above) also show positive values for most loci (mean = 0.50 ± 0.47), again consistent with balancing selection affecting the PAR genes.

Discussion

Our study of nucleotide polymorphisms in PAR genes was designed to test for the footprints of balancing selection,

Table 3 Haplotype tests using the estimated *R*/gene values

Locus	Map position	Excluded sequence numbers	Observed haplotype diversity[a]	Estimated *R*/gene	*K* test		*H* test	
					S	*P* value[b]	θ/gene	*P* value
E559	0	F2, 12	0.83	15.3	7	0.75, 0.74	2.89	0.32, 0.34
E200	0	M3, F2, 6	0.83	30.9	8	0.82, 0.45	2.13	0.42, 0.43
E241	0	M3, F1, F2	0.89	14	50	**0.017**, **0.012**	17.6	**0.004**, **0.002**
$E284_{PAR}$	0	M3, 8, 23, F1, 4, 8a, 10	0.98	30.2	102	**0.021**, **0.03**	34.5	**0.014**, **0.010**
E592	13	M11, 23, F2, 3b, 6, 9, 12	0.84	5.4	26	0.087, 0.057	12.0	**0.013**, **0.006**
E219	21	F6, 8, 9	0.87	0.001	43	0.114, 0.102	14.6	**0.035**, **0.027**

[a] Note that these analyses used only variable sites, so that the S and θ/gene values cannot be compared with those in other tables or the text.
[b] Probability of getting a haplotype diversity value lower than the observed value (2 × 1000 replicates). Bold font indicates *P*-values < 0.05.

which should generate LD with neutral variants in the region (Hudson and Kaplan 1988; Nordborg *et al.* 1996; Charlesworth *et al.* 1997), and could suggest the occurrence of polymorphisms caused by SA selection.

We observed high allelic or nucleotide diversity in the *S. latifolia* PAR genes studied, which seems to suggest some form of balancing selection (but, of course, not necessarily SA selection; see below for alternative possibilities). Indeed, only one gene, *E352*, shows the differentiation between males and females, and evidence for linkage disequilibrium with the MSY, that are predicted under the SA polymorphism hypothesis. However, the data are not currently conclusive, because the lack of allele frequency differences at the other loci may largely reflect the greater ease of detecting different alleles in the *E352* gene. For this gene, we observed high length variability in an intron, with alleles of different lengths readily distinguished by capillary electrophoresis; as explained above, the high variability prevented our obtaining sequences for the two alleles of many individuals. For the other genes, allele frequencies are less easily determined because, although we were able to obtain sequences, we find many different variants whose phase in the heterozygous plants is generally not known (Table S1); therefore we cannot directly compare haplotype frequencies in the two sexes, and all that can currently be said is that there are few sex differences in frequencies at individual nonsingleton sites. Given that different individuals within a species may differ in recombination rates (reviewed in Comeron *et al.* 2012), larger samples of alleles should be studied, once further PAR genes are identified. Ideally, sets of parents and offspring should be genotyped to ascertain the phase of variants in different PAR, to quantify LD across the PAR, which will reflect recombination in the species' recent history.

An important question is the extent to which the high nucleotide diversity in *S. latifolia* PAR genes can be explained simply by partial linkage to the MSY region, without the action of SA selection. As discussed in the Introduction, neutral models suggest that very close linkage is necessary. Clearly, recombination in the *S. latifolia* PAR should be further studied. This should also help resolve differences in the estimated genetic map distances from the MSY for the closely linked *E241* and *OPA* markers, which we find to be smaller than those estimated previously (Blavet *et al.* 2012). However, two genes, *E592* and *E219*, are at least 8 cM from the MSY (see table 3 of Bergero *et al.* 2013) and have high diversity (Table 1), and our tests gave no evidence of paralogues in the species' genome that could create a false appearance of high diversity. At such loose linkage, SA polymorphisms may appear no more plausible than in autosomal regions, where such polymorphisms are unlikely to be maintained (Patten *et al.* 2010; Jordan and Charlesworth 2012). A possibility that is consistent with high diversity at such distal locations is that linked SA polymorphisms are expected to develop LD with one another (Patten *et al.* 2010; Jordan and Charlesworth 2012). So far, this effect has been studied only for autosomal loci, but the effect on LD in the PAR of two or more such polymorphisms might be expected to be greater. This should be modeled in the future.

We have not attempted to explain our diversity results in terms of the direct effect of linkage to the MSY *vs.* the effect of selection maintaining associations with the male-determining factor, because the strength of selection is unknown. However, modeling Tajima's *D* could allow one to estimate the extra effect of selection, by modeling the expected value for a neutral PAR gene at recombination distances relevant for *S. latifolia*. This is beyond the scope of the present paper, because our polymorphism results from autosomal loci suggest a recent expansion of the species. The parameters of the species' recent demographic history that will affect Tajima's *D* must therefore also be estimated quantitatively, using autosomal polymorphisms. Given parameter estimates, one could conduct modeling to predict Tajima's *D* for PAR genes, using the inferred history and recombination distances.

An intriguing observation, consistent with the *S. latifolia* PAR tending to harbor high polymorphism levels, is the unexpectedly high X–Y divergence for one gene that has recently become fully sex linked, *E521* (see Bergero *et al.* 2013). There is no reason to think that this gene has a high mutation rate, as divergence from *S. vulgaris* is not out of line with that of other loci (Table 1). Such high divergence thus suggests that the region close to this locus was probably highly polymorphic before recombination between the Y and X was suppressed. If strong LD existed between the MSY and an *E521* haplotype associated with a male-benefit allele maintained polymorphic in the region for a long time, as expected under the SA selection hypothesis, suppression of recombination with the Y would link the male-advantage

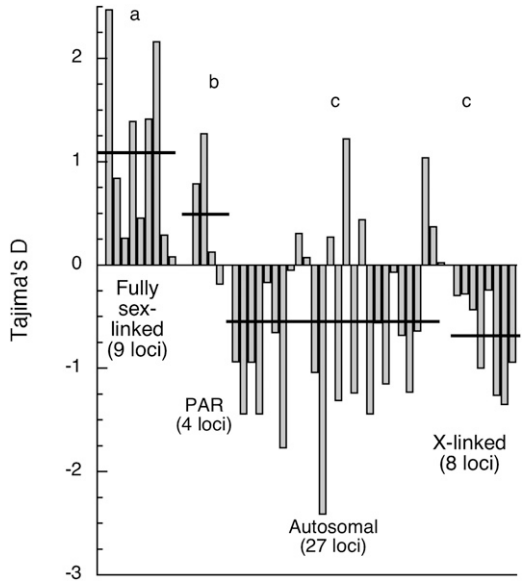

Figure 1 Tajima's *D* values for autosomal, fully sex-linked, and PAR genes. The values shown are for all site types (data from two new PAR genes, plus new analyses of genes studied by Qiu *et al.* (2010) and Muir *et al.* (2011), provided that there were at least 10 segregating sites). The lines across each of the sets of genes indicate the mean value for that set. The lowercase letters indicate sets of genes with significantly different mean values, using Mann–Whitney *U*-tests.

allele to the MSY, while the X chromosome population would lose this allele. This would convert some of the diversity in the population (*i.e.*, preexisting differences between haplotypes carrying the male-benefit alleles and those without them) into fixed Y–X differences. If such events have occurred during sex chromosome evolution, the times when recombination stopped may be overestimated for loci in recently formed strata, because sequence differences between alleles predate the event that caused recombination suppression.

In mammals, parts of the PAR experience an exceptionally high recombination rate, due to a requirement for a crossing over event in male meiosis in this physically small region, and this appears to cause a high mutation rate (Yi and Li 2005; Bussell *et al.* 2006). A high mutation rate could lead to high diversity at pseudoautosomal genes. However, this is unlikely in *S. latifolia*, because the region is unlikely to be physically very small, since the map length is large at least 20 cM in both sexes. Consistent with these observations, our HKA tests exclude a high mutation rate as the explanation for the high diversity of these genes, even though divergence from *S. vulgaris* is slightly higher for *S. latifolia* PAR genes than for fully X-linked genes, suggesting a possible elevated mutation rates in the PAR (Table 2).

Alternatives to SA selection

High diversity in PAR genes could be due to some form of balancing selection other than SA effects maintaining different alleles of these or other PAR loci. However, it would be surprising if a small genome region contained several loci with balancing selection. We cannot exclude the possibility that different alleles of PAR genes are maintained in different chromosome arrangements, for example a polymorphic inversion in the PAR. However, recently established inversion polymorphisms are unlikely to cause greatly elevated diversity, as has recently been confirmed empirically in *Drosophila melanogaster* (Corbett-Detig and Hartl 2012). Future work in *S. latifolia* should nevertheless test this possibility. The small region of visible pairing in meiotic preparations probably makes cytological testing for an inversion impossible, but genetic or physical mapping of PAR genes in different families could be informative. It is now becoming possible to find new PAR genes, and markers in them, through recent technical developments such as restriction-site-associated DNA (RAD) sequencing for genetic mapping (Baird *et al.* 2008).

We cannot exclude the possibility that the high PAR region sequence diversity is due to introgression from *S. dioica*. This might also explain the non-Mendelian segregation we observed in one F_2 family (see Bergero *et al.* 2013), since such ratios can reflect incompatibilities between species that cause loss of some zygote genotypes (reviewed in Salome *et al.* 2012). However, introgression seems unlikely to explain diversity as high as that observed. Raw divergence estimates between *S. latifolia* and *S. dioica*, with Jukes–Cantor correction (*e.g.*, Graur and Li 2000), average 0.036 for silent sites, based on 10 loci for which sequences have been published, or 0.023 when weighted by the numbers of sites analyzed (to downweight two loci with sequences covering fewer than 125 silent sites). Silent site diversity values within our sample of *S. latifolia* plants exceed 5% for the four PAR genes other than *E200*, whose diversity is probably underestimated (see above; the values in Table 1 are lower because all sites were used in the estimates). To generate such high diversity, the divergence from *S. dioica* sequences would have to be higher than observed, and introgression would have to be very frequent.

A final possibility is that the *S. latifolia* PAR was previously nonrecombining, but recombination recently started to occur in the region. However, this would require a long evolutionary period of complete Y linkage, to build up the large number of variants we find. The results in Bergero *et al.* (2013) contradict this scenario, since the genes in the regions carrying the currently partially sex-linked genes were clearly added to both the X and Y, after the evolution of the younger of the two previously known strata, and therefore the added genome regions must initially have been able to recombine. Moreover, the fact that the PAR genes have been added to both the Y and the X indicates that the added genome regions must initially have been able to recombine.

Polymorphism for SA genes is nevertheless only one possible cause of recombination suppression. Other possibilities are not exclusive. For instance, recombination may occasionally occur in largely nonrecombining PAR boundary regions, as occurs in humans (Cooke *et al.* 1985). Such a region should have a reduced effective population size, which could allow transposable elements to accumulate, potentially

to the extent that recombination is directly hindered. The reduction in effective population size need not be very great. For example, X chromosomes in mammals and plants have a lower recombination frequency than the autosomes, because X chromosomes recombine only in females, and, in mammals and papaya, this chromosome has a detectably higher repetitive content than the genome-wide average (reviewed in Bergero *et al.* 2008). Transposable element heterozygosity is known to reduce recombination in maize (Dooner and Martinez-Férez 1997; Fu *et al.* 2002; Dooner and He 2008), and transposable elements are abundant in *S. latifolia* (Cermak *et al.* 2008; Macas *et al.* 2008), whose haploid genome size is similar to that of maize (Grover *et al.* 2008).

Acknowledgments

We thank GenePool, Edinburgh for the Sanger sequencing and capillary electrophoreses, and the Biotechnology and Biological Sciences Research Council for funding.

Note added in proof: See Bergero *et al*. 2013 (pp. 673–686) in this issue, for a related work.

Literature Cited

Baird, N., P. Etter, T. Atwood, M. Currey, A. Shiver *et al.*, 2008 Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3: e3376.

Bergero, R., and D. Charlesworth, 2011 Preservation of the Y transcriptome in a 10MY old plant sex chromosome system. Curr. Biol. 21: 1470–1474.

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., A. Forrest, and D. Charlesworth, 2008 Active miniature transposons from a plant genome and its nonrecombining Y chromosome. Genetics 178: 1085–1092.

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

Blavet, N., H. Blavet, R. Cegan, N. Zemp, J. Zdanska *et al.*, 2012 Comparative analysis of a plant pseudoautosomal region (PAR) in *Silene latifolia* with the corresponding *S. vulgaris* autosome. BMC Genomics 13: 226.

Bonduriansky, R., and S. F. Chenoweth, 2009 Intralocus sexual conflict. Trends Ecol. Evol. 25: 280–288.

Bussell, J. J., N. M. Pearson, R. Kanda, D. A. Filatov, and B. T. Lahn, 2006 Human polymorphism and human-chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. Gene 368: 94–100.

Calafell, F., F. Roubinet, A. Ramírez-Soriano, N. Saitou, J. Bertranpetit *et al.*, 2008 Evolutionary dynamics of the human ABO gene. Hum. Genet. 124: 123–135.

Cermak, T., Z. Kubat, R. Hobza, A. Koblizkova, A. Widmer *et al.*, 2008 Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes. Chromosome Res. 16: 961–976.

Charlesworth, B., I. Mori, and D. Charlesworth, 1985 Genetic variation in recombination in Drosophila. 3.3. Regional effects on crossing over and effects on non-disjunction. Heredity 55: 209–221.

Charlesworth, B., M. Nordborg, and D. Charlesworth, 1997 The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided inbreeding and outcrossing populations. Genet. Res. 70: 155–174.

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.

Chibalina, M., and D. Filatov, 2011 Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr. Biol. 21: 1475–1479.

Comeron, J., R. Ratnappan, and S. Bailin, 2012 The many landscapes of recombination in *Drosophila melanogaster*. PLoS Genet. 8: e1002905.

Cooke, H., W. Brown, and G. Rappold, 1985 Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. Nature 317: 687–692.

Corbett-Detig, R. B., and D. L. Hartl, 2012 Population genomics of inversion polymorphisms in *Drosophila melanogaster*. PLoS Genet. 8: e1003056.

Delcourt, M., M. W. Blows, and H. D. Rundle, 2009 Sexually antagonistic genetic variance for fitness in an ancestral and a novel environment. Proc. Biol. Sci. 276: 2009–2014.

Delph, L., A. Arntz, C. Scotti-Saintagne, and I. Scotti, 2010 The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. Evolution 64: 2873–2886.

Delph, L. F., and C. R. Herlihy, 2012 Sexual, fecundity, and viability selection on flower size and number in a sexually dimorphic plant. Evolution 66: 1154–1166.

DePaulis, F., and M. Veuille, 1998 Neutrality tests based on the distribution of haplotypes under an infinite sites model. Mol. Biol. Evol. 15: 1788–1790.

Dooner, H., and L. He, 2008 Maize genome structure variation: interplay between retrotransposon polymorphisms and genic recombination. Plant Cell 20: 249–258.

Dooner, H. K., and I. M. Martinez-Férez, 1997 Recombination occurs uniformly in the *bronze* gene, a recombination hotspot in the maize genome. Plant Cell 9: 1633–1646.

Ebinuma, H., 1987 Selective recombination system in *Bombyx mori*. 1. chromosome specificity of the modification effect. Genetics 117: 521–531.

Ellegren, H., 2011 Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat. Rev. Genet. 12: 157–166.

Eronen, L., F. Geerts, and H. Toivonen, 2006 HaploRec: efficient and accurate large-scale reconstruction of haplotypes. BMC Bioinformatics 7: 542.

Filatov, D. A., F. Monéger, I. Negrutiu, and D. Charlesworth, 2000 Evolution of a plant Y-chromosome: variability in a Y-linked gene of *Silene latifolia*. Nature 404: 388–390.

Filatov, D. A., V. Laporte, C. Vitte, and D. Charlesworth, 2001 DNA diversity in sex linked and autosomal genes of the plant species *Silene latifolia* and *S. dioica*. Mol. Biol. Evol. 18: 1442–1454.

Fu, H., Z. Zheng, and H. K. Dooner, 2002 Recombination rates between adjacent genic and retrotransposon regions in maize vary by 2 orders of magnitude. Proc. Natl. Acad. Sci. USA 99: 1082–1087.

Gordon, S. P., A. López-Sepulcre, and D. N. Reznick, 2012 Predation-associated differences in sex-linkage of wild guppy coloration. Evolution 66: 912–918.

Graur, D., and W.-H. Li, 2000 *Fundamentals of Molecular Evolution*. Sinauer Associates, Sunderland, Massachusetts.

Grover, C., J. Hawkins, and J. Wendel, 2008 Phylogenetic insights into the pace and pattern of plant genome size evolution, pp. 57–68 in *Plant Genomes*, edited by J.-N. Volff Karger, Basel, Switzerland.

Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics 120: 831–840.

Hudson, R. R., D. D. Boos, and N. L. Kaplan, 1992 A statistical test for detecting geographic subdivision. Mol. Biol. Evol. 9: 138–151.

Iwase, M., Y. Satta, Y. Hirai, and N. Takahata, 2010 Frequent gene conversion events between the X and Y homologous chromosomal regions in primates. BMC Evol. Biol. 10: 225.
Jordan, C., and D. Charlesworth, 2012 The potential for sexually antagonistic polymorphism in different genome regions. Evolution 66: 505–516.
Kelly, J. K., 1997 A test of neutrality based on interlocus associations. Genetics 146: 1197–1206.
Kirkpatrick, M., R. Guerrero, and S. Scarpino, 2010 Patterns of neutral genetic variation on recombining sex chromosomes. Genetics 184: 1141–1152.
Kitano, J., J. A. Ross, S. Mori, M. Kume, F. C. Jones *et al.*, 2009 A role for a neo-sex chromosome in stickleback speciation. Nature 461: 1079–1083.
Lahn, B. T., and D. C. Page, 1999 Four evolutionary strata on the human X chromosome. Science 286: 964–967.
Laporte, V., D. A. Filatov, E. Kamau, and D. Charlesworth, 2005 Indirect evidence from DNA sequence diversity for genetic degeneration of Y-chromosome in dioecious species of the plant *Silene*: the *SlY4/SlX4* and *DD44*-X/*DD44*-Y gene pairs. J. Evol. Biol. 18: 337–347.
Lawson-Handley, L. J., H. Ceplitis, and H. Ellegren, 2004 Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics 167: 367–376.
Lemaitre, C., M. D. V. 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.
Librador, P., and J. Rozas, 2009 DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.
Lindholm, A., and F. Breden, 2002 Sex chromosomes and sexual selection in Poeciliid fishes. Am. Nat. 160: S214–S224.
Liu, Z., P. H. Moore, H. Ma, C. M. Ackerman, M. Ragiba *et al.*, 2004 A primitive Y chromosome in Papaya marks the beginning of sex chromosome evolution. Nature 427: 348–352.
Macas, J., E. Kejnovsky, P. Neumann, P. Nova, A. Koblizkova *et al.*, 2008 Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. PLoS ONE 6: e27335.
Mank, J. E., and H. Ellegren, 2009 Sex-linkage of sexually antagonistic genes is predicted by female, but not male, effects in birds. Evolution 63: 1464–1472.
Mank, J. E., L. Hultin-Rosenberg, M. Zwahlen, and H. Ellegren, 2008 Pleiotropic constraint hampers the resolution of sexual antagonism in vertebrate gene expression. Am. Nat. 171: 35–43.
Matsubara, K., H. Tarui, M. Toriba, K. Yamada, C. Nishida-Umehara *et al.*, 2006 Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. Proc. Natl. Acad. Sci. USA 103: 18190–18195.
Matsunaga, S., E. Isono, E. Kejnovsky, B. Vyskot, S. Kawano *et al.*, 2003 Duplicative transfer of a MADS box gene to a plant Y chromosome. Mol. Biol. Evol. 20: 1062–1069.
McDaniel, S. F., J. H. Willis, and A. J. Shaw, 2007 A linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss *Ceratodon purpureus*. Genetics 176: 2489–2500.
McVean, G. A. T., P. Awadalla, and P. Fearnhead, 2002 A coalescent-based method for detecting and estimating recombination from gene sequences. Genetics 160: 1231–1241.
Muir, G., C. J. Dixon, A. L. Harper, and D. A. Filatov, 2011 Dynamics of drift, gene flow and selection during speciation in *Silene*. Evolution 66: 1447–1458.
Nam, K., and H. Ellegren, 2008 Scrambled eggs: the chicken (*Gallus gallus*) Z chromsome contains at least three non-linear evolutionary strata. Genetics 180: 1131–1136.
Nishida-Umehara, C., Y. Tsuda, J. Ishijima, J. Ando, A. Fujiwara *et al.*, 2007 The molecular basis of chromosome orthologies and sex chromosomal differentiation in palaeognathous birds. Chromosome Res. 15: 721–734.
Nordborg, M., B. Charlesworth, and D. Charlesworth, 1996 Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. Biol. Sci. 163: 1033–1039.
Otto, S. P., J. R. Pannell, C. L. Peichel, T. Ashman, D. Charlesworth *et al.*, 2011 About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet. 27: 358–367.
Patten, M. M., D. Haig, and F. Úbeda, 2010 Fitness variation due to sexual antagonism and linkage disequilibrium. Evolution 64: 3638–3642.
Perkins, D. D., and M. Bojko, 1992 The basis of decreased recombination in certain outcrosses of *Neurospora crassa*. Genome 35: 503–509.
Qiu, S., R. Bergero, A. Forrest, V. B. Kaiser, and D. Charlesworth, 2010 Nucleotide diversity in *Silene latifoli*a autosomal and sex-linked genes. Proc. Biol. Sci. 277: 3283–3290.
Rice, W. R., 1987 The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex-chromosomes. Evolution 41: 911–914.
Roberts, R. B., J. R. Ser, and T. D. Kocher, 2009 Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science 326: 998–1001.
Rosser, Z. H., P. Balaresque, and M. A. Jobling, 2009 Gene conversion between the X chromosome and the male-specific region of the Y chromosome at a translocation hotspot. Am. J. Hum. Genet. 851: 130–134.
Salome, P., K. Bomblies, J. Fitz, R. Laitinen, N. Warthmann *et al.*, 2012 The recombination landscape in *Arabidopsis thaliana* F2 populations. Heredity 108: 447–455.
Ségurel, L., E. E. Thompson, T. Flutre, J. Lovstad, A. Venkat *et al.*, 2012 The ABO blood group is a trans-species polymorphism in primates. Proc. Natl. Acad. Sci. USA 109: 18493–18498.
Skaletsky, H., T. Kuroda-Kawaguchi, P. J. Minx, H. S. Cordum, L. Hillier *et al.*, 2003 The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423: 825–837.
Slattery, J. P., L. Sanner-Wachter, and S. J. O'Brien, 2000 Novel gene conversion between X-Y homologues located in the nonrecombining region of the Y chromosome in Felidae (Mammalia). Proc. Natl. Acad. Sci. USA 97: 5307–5312.
Spigler, R., K. Lewers, D. Main, and T.-L. Ashman, 2008 Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. Heredity 101: 507–517.
Stulp, G., B. Kuijper, A. P. Buunk, T. V. Pollet, and S. Verhulst, 2012 Intralocus sexual conflict over human height. Biol. Lett. 8: 976–978.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis. Genetics 123: 585–595.
Trombetta, B., F. Cruciani, P. Underhill, D. Sellitto, and R. Scozzari, 2010 Footprints of X-to-Y gene conversion in recent human evolution. Mol. Biol. Evol. 27: 714–725.
Úbeda, F., D. Haig, and M. M. Patten, 2011 Stable linkage disequilibrium owing to sexual antagonism. Proc. Biol. Sci. 278: 855–862.
Wang, J., J. Na, Q. Yu, A. R. Gschwend, J. Han *et al.*, 2012 Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl. Acad. Sci. USA 109: 13710–13715.
Wright, S. I., and B. Charlesworth, 2004 The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168: 1071–1076.
Yi, S., and W.-H. Li, 2005 Molecular evolution of recombination hotspots and highly recombining pseudoautosomal regions in Hominoids. Mol. Biol. Evol. 22: 1223–1230.

Communicating editor: S. I. Wright

GENETICS

Supporting Information
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152397/-/DC1

Testing for the Footprint of Sexually Antagonistic Polymorphisms in the Pseudoautosomal Region of a Plant Sex Chromosome Pair

Suo Qiu, Roberta Bergero, and Deborah Charlesworth

DOI: 10.1534/genetics.113.152397

File S1

Tests for paralogues of the PAR genes and for "null alleles"

For two genes, *E241* and *E284*, Neighbour-Joining tree analyses identified distinctive group of sequences suggesting potential paralogues. These trees are not shown, because these are recombining sequences and therefore estimating phylogenetic trees is not an appropriate analysis, and can be used only to help visualise the data and suggest suitable formal analyses.

For *E241*, we analysed the segregation of a potentially paralogous set of sequences in a family E2008-5, whose maternal parent (K2005-7/4, see Bergero *et al.* 2013) has both sequence types; The two sequence types segregated as expected for alleles, so we conclude that there is no paralogous copy.

For *E284*, a group of 11 sequences form a distinctive haplotype that could represent a duplicate copy. PCR reactions with primers specific for this sequence type yielded an amplicon of the expected sequence in all 24 male plants tested from natural populations, supporting the presence of a paralogous *E284* copy. This does not affect our genetic results, since the two sequences that amplified in the parents of our families, using our genotyping primers, segregate as alleles and provided strong evidence for a PAR location; we therefore denote this gene by $E284_{PAR}$. Our diversity study used only $E284_{PAR}$ sequences. We attempted to map the putative paralogue by genotyping family G2008-3 (see Bergero *et al.* 2013) using primers specific for its sequence; the paternal plant is heterozygous for three variants in this paralogous sequence (in total, this plant has 4 different *E284* sequences, 2 assigned to the $E284_{PAR}$ gene, and 2 to the paralogue). Many female progeny inherited the paralogue from the paternal plant, showing that it is not a duplication onto the Y, but the small size of the family did not allow us to determine definitively whether it is autosomal or pseudo-autosomal.

We also checked PAR variants in our sample of multiple individuals from natural populations, reasoning that paralogous copies would be found in many individuals, whereas divergent alleles would be found only in

certain individuals. For these tests, we used a larger sample, including all 24 male plants surveyed in our previous study of natural populations (Qiu *et al.* 2010).

We also used the occurrence of homozygous sequences from the PAR genes in our sample of males as a further indication that the gene is in a region that recombines with the sex-determining region (since, for fully sex linked genes, all males should be heterozygous at sites with fixed Y-linked variants, and this is indeed found for fully sex-linked loci, see Results section). This test excludes the possibility that Y-alleles are present that fail to amplify with the primers used, because such sequences would behave as null alleles and create the appearance of homozygous genotypes for all males. Because the presence of undetectable alleles at partially sex-linked loci could lead to our under-estimating nucleotide diversity, we tested for such null alleles using genotype frequencies in the population samples where an excess of homozygotes was apparent.

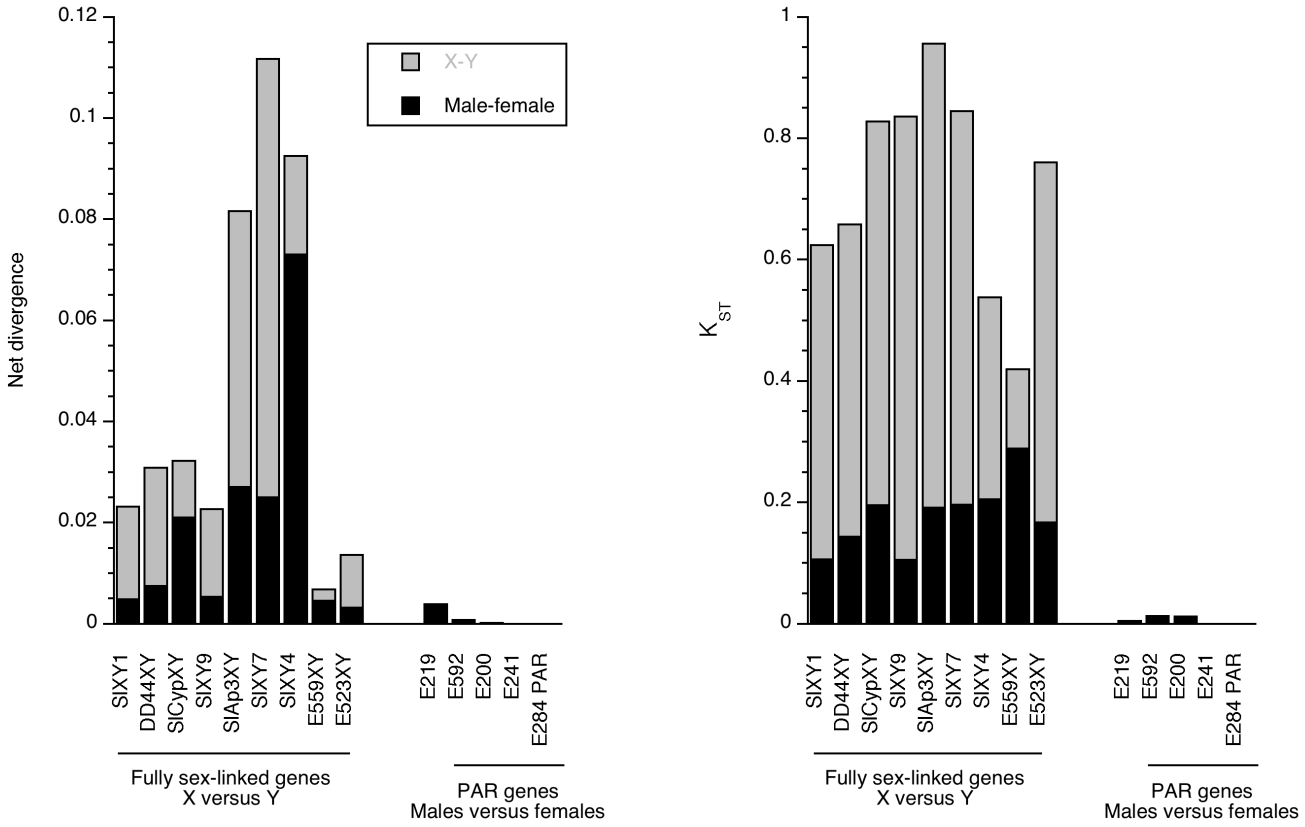

Figure S1 Net divergence between sequences from males and females, and K_{ST}.

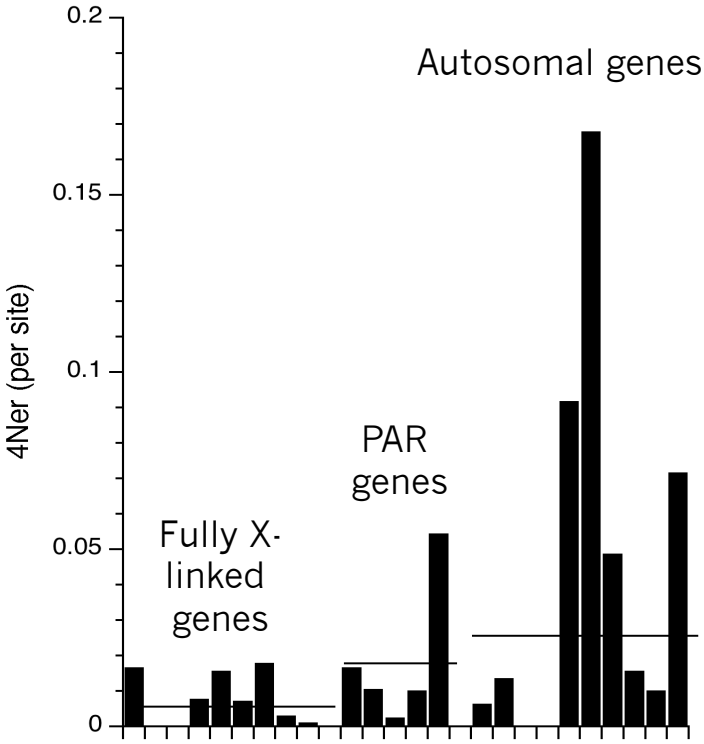

Figure S2 Sex averaged recombination rates for autosomal, fully X-linked and PAR genes, estimated from the sequences, using LDhat software.

Table S1 *S. latifolia* individuals obtained from field-collected seeds across Europe, and the sequences obtained.

Available for download at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.113.152397/-/DC1.

Table S2 Primers used for the molecular diversity analyses of PAR genes and the newly discovered fully sex-linked E559 gene in this study.

Genes	Primers (forward)	Primers (reverse)
E200	E200-f8: AAGAGAATCGACATGACTTTCAGGC	E200-ri3: GGGTTTATACGTCATTACGATTCACC
E219	E219-f3: TCGACGGTGTTTCATCAATCTGTGC	E219-r3: CCCTATTCTGAAAACCTTGCCAGC
E241	E241-f3: AAGAGTGGAAAGTAAAGCGCGAAG	E241-r2: GCGACTATTATCTTGGCATTTTTGTC
E284	E284-f2: GTTGTTGCTTTATTGGTTTTTGGTCC	E284-r3: ACTCAAAAATCATAGCAGCACAAGG
E559	E559-f2: ACATGGAGAATACTTCGAAGTGCACC	E559-r: ACTTGTTTCCTCAAGGATGACACC
E592	E592-f2: GAGATCATACAAGCATCAGACGGAGC	E592-r2: AATCATTCCTCTGGCTTTCAGCAAAGC

Table S3 Summary of LD analyses, excluding all sites with > 2 variants (and also excluding one female plant F13 plant in Table S1 whose sequences are incomplete). The phases of variants were inferred using HaploRec, using a minimum allele frequency for variants of 10%. The loci are ordered according to our estimated recombination distances from the PAR boundary.

Gene	Number of sequences analyzed	Sequence numbers excluded from analysis	Number of variants	Zns [1]	Number of significant Fisher's exact tests/Number of comparisons (after Bonferroni correction)	Rmin [2]
E559	40	F2, 12 (or also excluding M3, M18)	7	0.277	6/21 = 0.29	3
E200	38	M3, F2, 6	8	0.314	5/28 = 0.18	2
E241	38	M3, F1, F2	50	0.193	158/1225 = 0.13	9
$E284_{PAR}$	31	M3, 8, 23, F1, 4, 8a, 10	102	0.167	181/5151 = 0.04	19
E592	31	M11, 23, F2, 3b, 6, 9, 12	276	0.414	135/351 = 0.38	5
E219	38	F6, 8, 9	43	0.565	461/903 = 0.51	7
First 4 genes	25	M3, 8, 23, F1, 2, 4, 6, 8a, 10, 12	263	0.139	57/3240= 0.018	19
All genes	20	As row above, plus M11, F3b, 9	224	0.139	— (too many variants to test)	38

[1] The linkage disequilibrium measure of JK Kelly (1997 A test of neutrality based on interlocus associations. Genetics 146, 1197-1206).

[2] The minimum number of recombination events (Hudson RR, Kaplan NL, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147-164).