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.

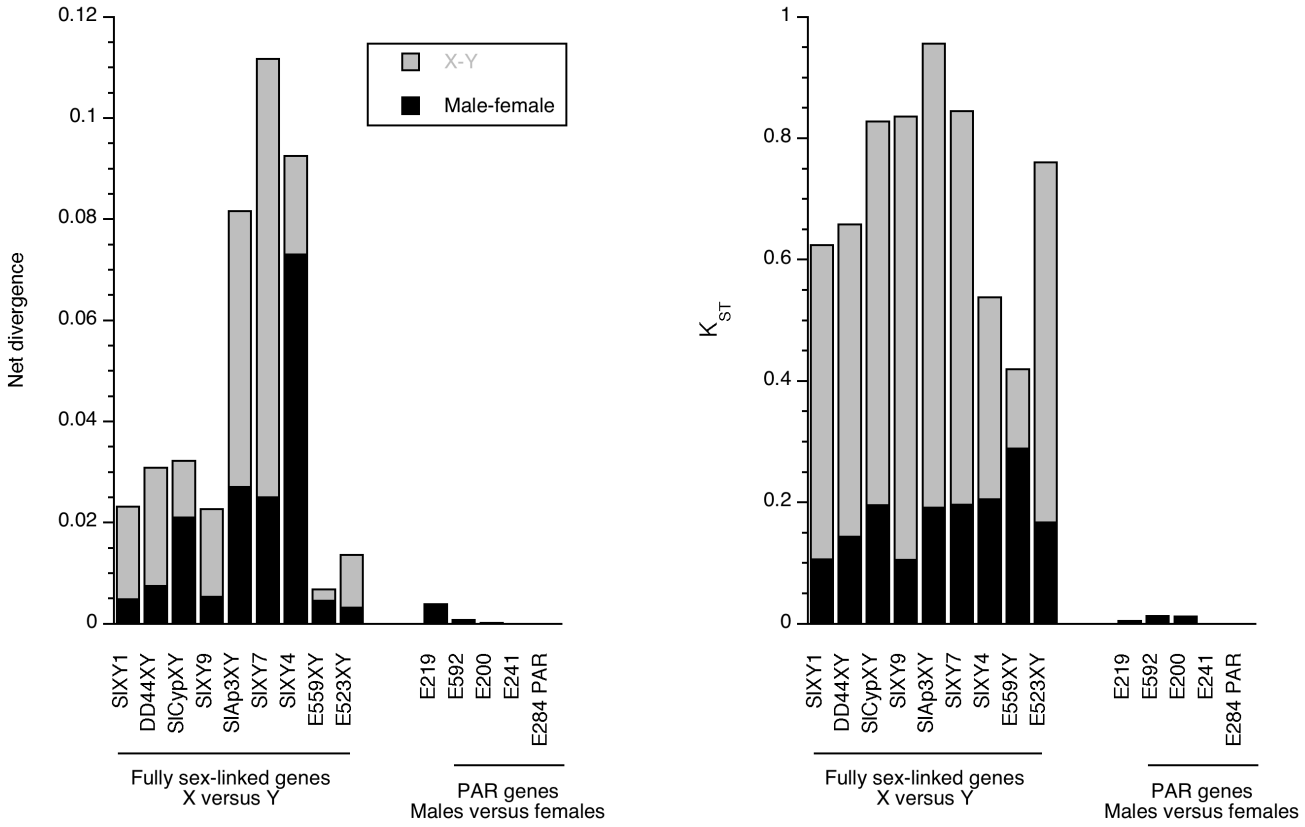

Figure S1 Net divergence between sequences from males and females, and K_{ST}.

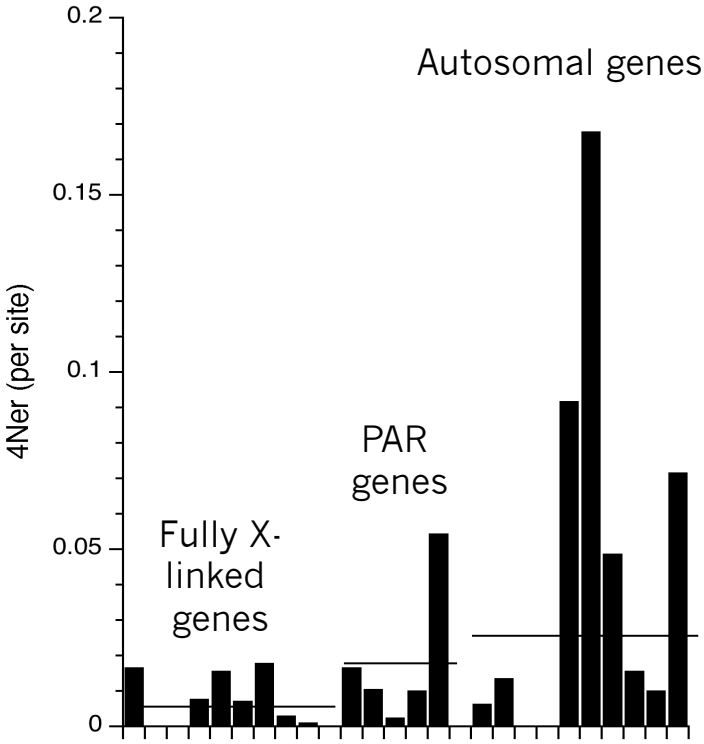

Figure S2 Sex averaged recombination rates for autosomal, fully X-linked and PAR genes, estimated from the sequences, using LDhat software.

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).