

## 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<sub>PAR</sub>*. Our diversity study used only *E284<sub>PAR</sub>* 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<sub>PAR</sub>* 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.