Genetic Diversity in the UV Sex Chromosomes of the Brown Alga *Ectocarpus*

K. Avia[1,2,*,†,] A.P. Lipinska[1,†], L. Mignerot[1], A.E. Montecinos[2,3], M. Jamy[1], S. Ahmed[1], M. Valero[2], A.F. Peters[4], J.M. Cock[1], D. Roze[2], S.M. Coelho[1]

Supplementary Methods

1. Genetic map of E. siliculosus

1.1.   Genetic mapping segregating population

A diploid sporophyte (Ec236) was generated by crossing two compatible *E. siliculosus* strains from a Naples population (EA1 and RB1) (Table S1). From this sporophyte, 152 haploid gametophytes were isolated, arising each from a unique meiotic event [1,2]. The sex of each individual was determined using sex-specific PCR markers [3].

1.2.   DNA extraction, RAD sequencing and analysis

DNA was extracted from the progeny of Ec236 and also from the parental strains RB1 and EA1 using the NucleoSpin 96 Plant II kit (Macherey-Nagel GmbH & Co. KG, Germany), according to the manufacturer's instructions. DNA quality was checked on agarose gels and the quantity measured by PicoGreen® (Fisher Scientific).

Double digest RAD library (ddRAD-seq) was carried out following Peterson*, et al.* [4]. The genomic DNA was digested with the restriction enzymes HhaI and SphI (New England Biolabs, https://www.neb.com/). Cleaning, barcode ligation, PCR and size selection were performed as described in Avia*, et al.* [5]. The library was then sequenced on an Illumina Hiseq 2500 platform (Rapid Run Mode).

The ddRAD-seq sequencing data was analyzed using the Stacks pipeline (version 1.47, [6] as described in [5]. Final reads were trimmed to 100 bp. A *de novo*_map approach was used to call genotypes of the progeny based on those of their parents.

1.3.   Construction of the genetic map of *E. siliculosus*

SNP genotypes were exported from the Stacks pipeline in a vcf format. As haploid samples were used here, heterozygous genotypes were not expected. These could be due to sequencing errors or repetitive sequences. We filtered out heterozygous sites by removing all sites with more than 10% of samples being heterozygous and setting the remaining to missing. Monomorphic sites were also filtered out. Sites with a very high coverage, suggesting paralogs were also filtered out. To produce the final vcf file, we retained only sites with less than 10% of missing and removed all samples with more than 25% of missing data.

Lep-Map version3 (Lep-Map3, [7]) was used to build the genetic map. Lep-Map3 software allows mapping high-throughput whole genome sequencing datasets, even with

low-coverage through an efficient algorithm that reduce data filtering and curation, yielding more markers in the maps. The final vcf file and a pedigree file of our design were supplied to the different module of the software to form linkage groups and order markers within the groups. Optimal order was determined based on the best likelihood of 10 runs. The final order of the markers and their phased genotypes were saved for plotting and QTL mapping.

2. Alternative method to analyze neutral diversity in the non-recombining versus recombining regions

The results obtained with ddRAD-seq genome mapping and SNP calling using samtools mpileup method were confirmed by employing an alternative approach, based on using the Stacks pipeline [6] to analyze the ddRAD-seq data. The version 2 (beta8) of the software was used to assemble raw cleaned reads *de novo* into stacks. A catalog was built with the three samples from each population having the largest number of reads. Individuals stacks of every sample were then matched to the catalog to call SNPs. In order to have a correspondence between this *de novo* genotyping and the reference genome (*Ectocarpus* 7)-based analysis in the first method described above, consensus sequences from the catalog were mapped onto the reference *Ectocarpus* 7 genome and the coordinates of the uniquely mapped sequences were integrated into the *de novo* analysis (as in [8,9]). Final results were exported as a vcf file. The main difference between the principal method and this method is that, in this method, only bi-allelic variant positions were called with the Stacks pipeline.

References

1. Coelho, S.M.; Scornet, D.; Rousvoal, S.; Peters, N.; Dartevelle, L.; Peters, A.F.; Cock, J.M., Genetic crosses between *ectocarpus* strains. *Cold Spring Harbor Protocols* **2012**, *2012*, pdb. prot067942.
2. Coelho, S.M.; Scornet, D.; Rousvoal, S.; Peters, N.T.; Dartevelle, L.; Peters, A.F.; Cock, J.M., How to cultivate ectocarpus. *Cold Spring Harbor Protocols* **2012**, *2012*, pdb. prot067934.
3. Lipinska, A.P.; Ahmed, S.; Peters, A.F.; Faugeron, S.; Cock, J.M.; Coelho, S.M., Development of pcr-based markers to determine the sex of kelps. *Plos One* **2015**, *10*.
4. Peterson, B.K.; Weber, J.N.; Kay, E.H.; Fisher, H.S.; Hoekstra, H.E., Double digest radseq: An inexpensive method for de novo snp discovery and genotyping in model and non-model species. *PloS one* **2012**, *7*, e37135.
5. Avia, K.; Coelho, S.M.; Montecinos, G.J.; Cormier, A.; Lerck, F.; Mauger, S.; Faugeron, S.; Valero, M.; Cock, J.M.; Boudry, P., High-density genetic map and identification of qtls for responses to temperature and salinity stresses in the model brown alga *ectocarpus*. *Scientific Reports* **2017**, *7*, 43241.
6. Catchen, J.; Hohenlohe, P.A.; Bassham, S.; Amores, A.; Cresko, W.A., Stacks: An analysis tool set for population genomics. *Molecular Ecology* **2013**, *22*, 3124-3140.
7. Rastas, P., Lep-map3: Robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* **2017**, *33*, 3726-3732.
8. Paris, J.R.; Stevens, J.R.; Catchen, J.M., Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution* **2017**, *8*, 1360-1373.
9. Rochette, N.C.; Catchen, J.M., Deriving genotypes from rad-seq short-read data using stacks. *nature protocols* **2017**, *12*, 2640.
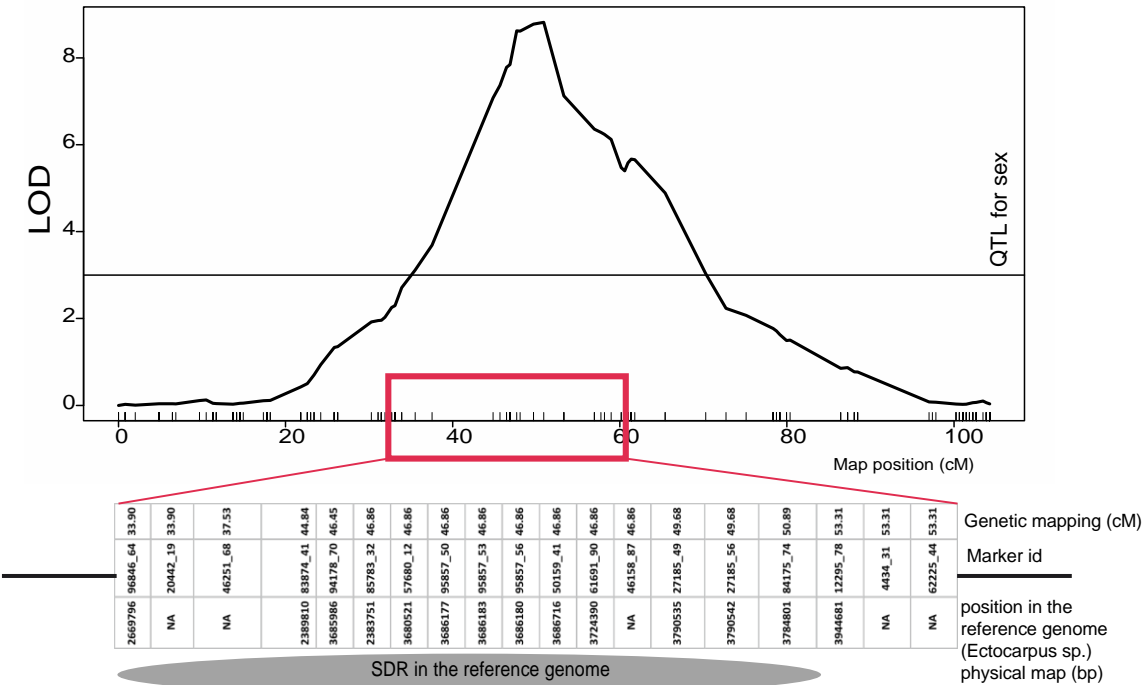
Supplementary Figures

Figure S1. Mapping of the sex-determining region in *Ectocarpus siliculosus*. Sex-determining region was assigned to linkage group 2 that contained 362 markers over 104.3 cM. A QTL peak corresponding to sexual identity (male or female) was detected at 50.9 cM. Positions of the E. siliculosus sex-markers on the reference genome (Ectocarpus sp.) physical map are shown below demonstrating overall colocalisation of the SDR boundary between the two species.
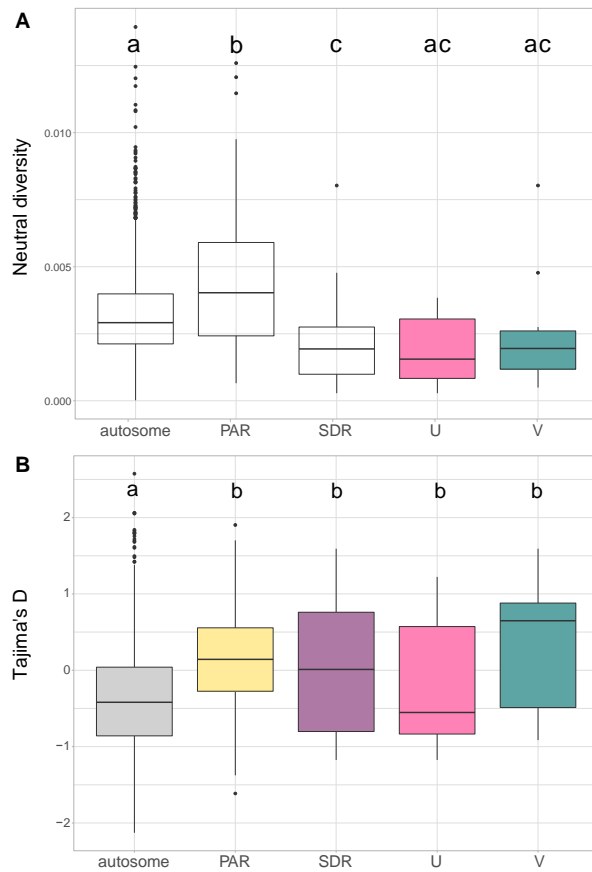
Figure S2. Population genetics statistics for the three *Ectocarpus siliculosus* genomic compartments: autosomes, PAR and SDR. A) Boxplots of neutral diversity (π), calculated in 1kb windows without overlap. **B)** Boxplots of Tajima's D, calculated in 1kb windows without overlap. Diversity and Tajima's D were calculated for the U and V separately and plotted separately or combined as SDR. Letters above the boxplot denote significant differences (Wilcoxon test, p<0.05).
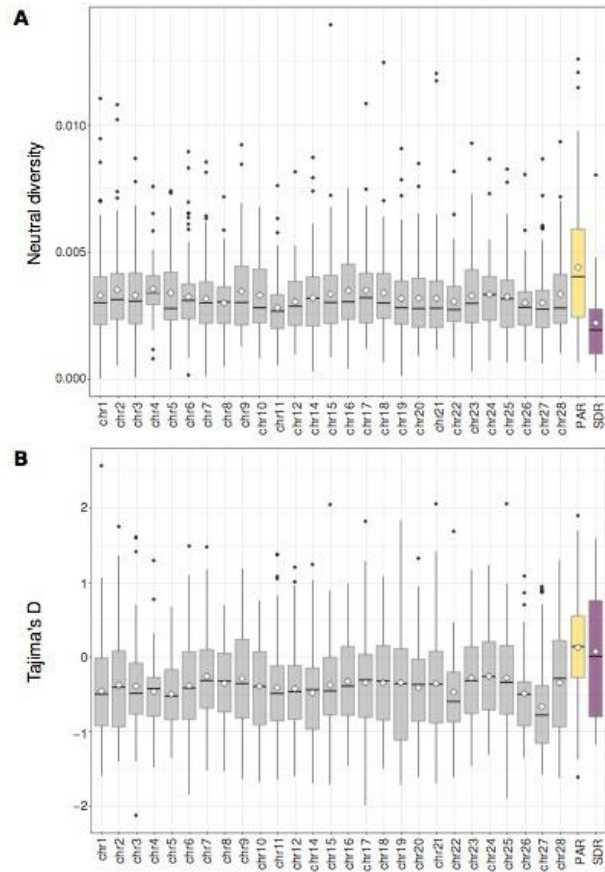
Figure S3. Population genetics statistics for the individual autosomes, PAR and SDR in *Ectocarpus siliculosus.* **A)** Boxplots of neutral diversity (π), calculated in 1kb windows without overlap. **B)** Boxplots of Tajima's D, calculated in 1kb windows without overlap. The mean values are represented by the diamond shape.
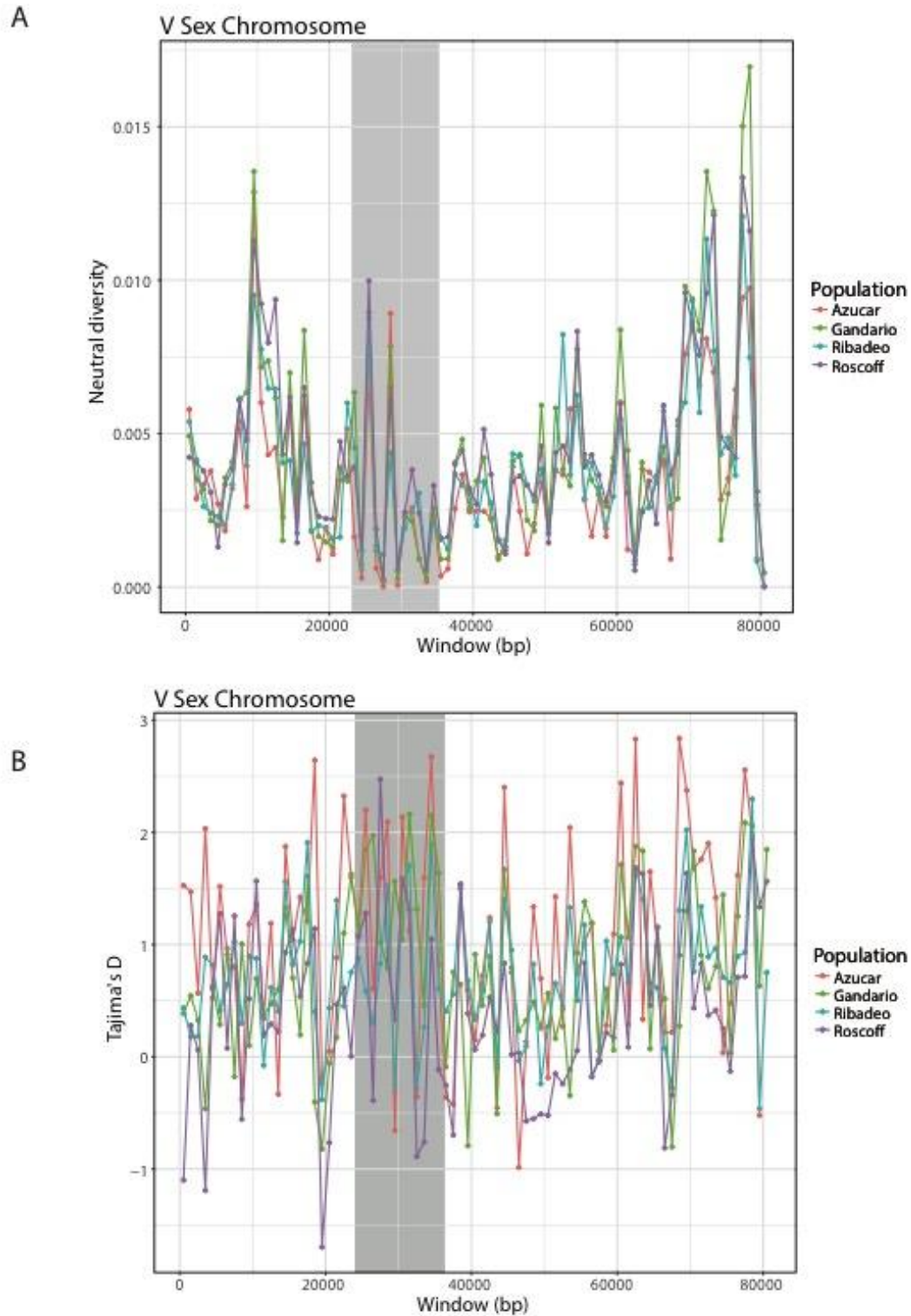
Figure S4. Comparison of neutral diversity patterns along the sex chromosome of *Ectocarpus siliculosus* between the four study populations. A) Sliding window analysis of neutral diversity (π) per 1kb non-overlapping windows along the concatenated ddRAD sequences of the sex chromosome. **B)** Sliding window analysis of Tajima's D per 1kb window along the concatenated sex chromosome sequences. The position of the sex-determining region is shaded in gray.
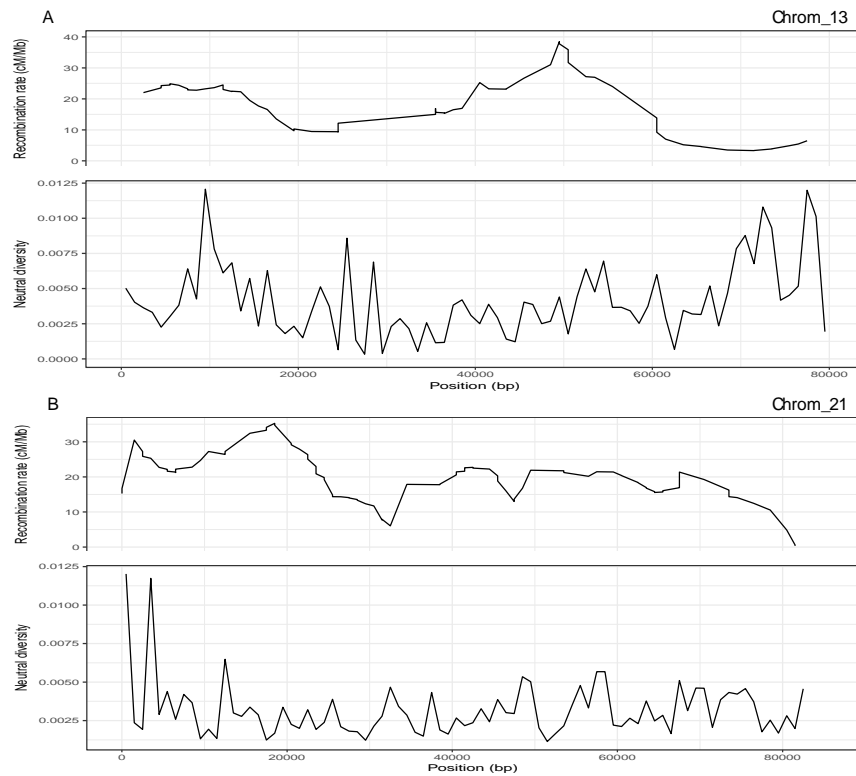
Figure S5. The landscape of neutral diversity (π) and recombination rates along the sex chromosome and autosomes in *Ectocarpus siliculosus*. The sex chromosome (chr13) and an autosome of similar size (chr21) are shown. The position of the sex-determining region is shaded in gray.
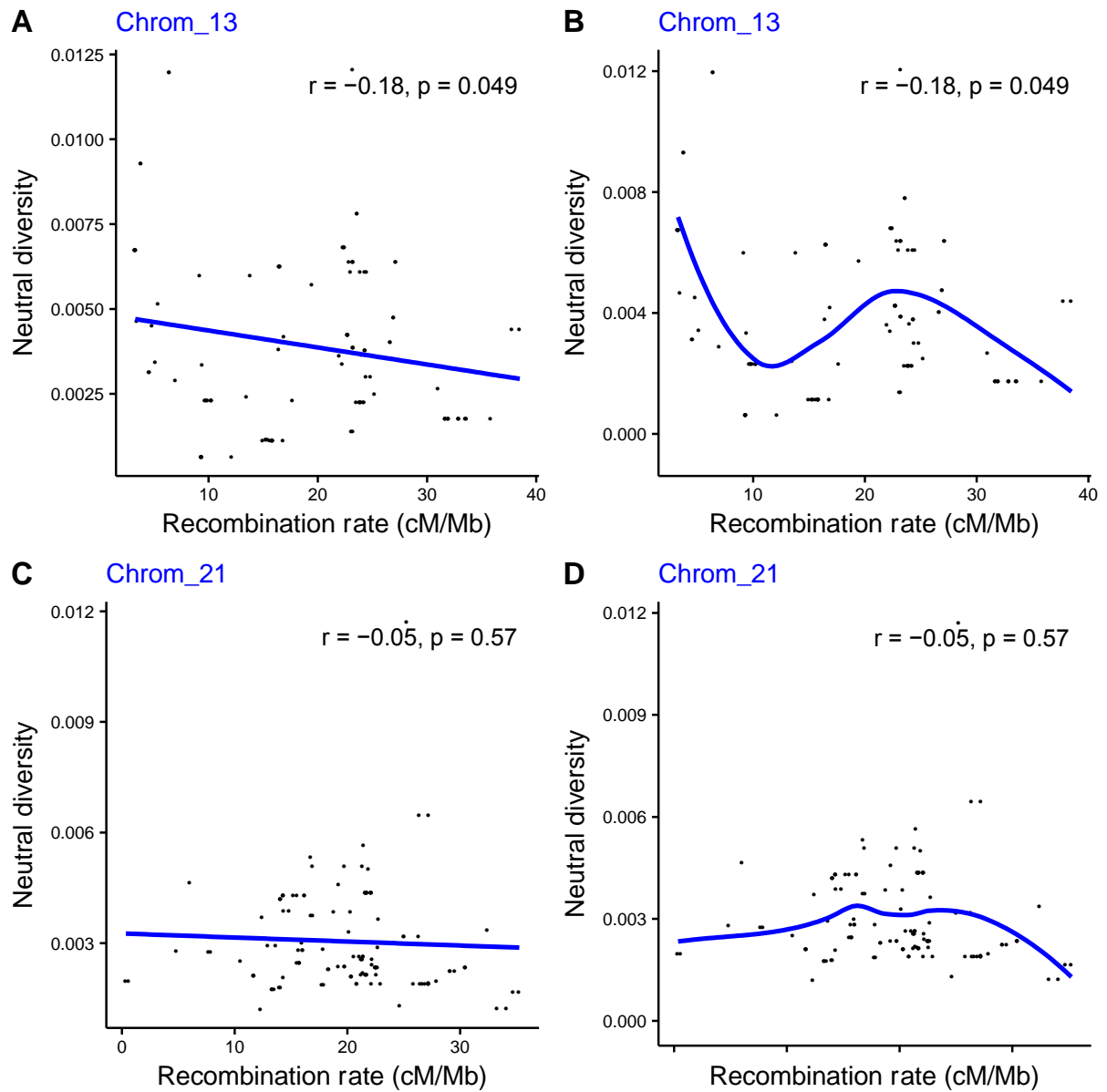
Figure S6. Correlation between neutral diversity and recombination rate. A and B: Correlation for the sex chromosome (chromosome 13) with linear regression line and local regression fitting respectively. C and D: Correlation for one autosome (chromosome 21) with linear regression line and local regression fitting respectively.

Supplementary Methods (see Excel File)

Table S1. Samples used for genetic mapping of the sex locus of *E. siliculosus.*

Table S2. Natural populations used for ddRAD sequencing and ddRADseq data summary.

Table S3. Samples used in the analysis of the population genetics statistics in coding regions.

Table S4. PCR primers used for amplification of the PAR and SDR genes.

Table S5. Summary statistics of the VCF files.