A Haploid System of Sex Determination in the Brown Alga *Ectocarpus* sp.

Sophia Ahmed, J. Mark Cock, Eugenie Pessia, Remy Luthringer, Alexandre Cormier, Marine Robuchon, Lieven Sterck, Akira F. Peters, Simon M. Dittami, Erwan Corre, Myriam Valero, Jean-Marc Aury, Denis Roze, Yves Van de Peer, John Bothwell, Gabriel A.B. Marais, and Susana M. Coelho

SUPPLEMENTARY FIGURES

Figure S1. *Ectocarpus* strain pedigree and SDR recombination analysis. Related to Figure 1.

(A) Pedigree of the *Ectocarpus* strains used in this study. SP, diploid sporophyte; GA, gametophyte; m, male; f, female; i, *immediate upright* mutant; ii, homozygous *immediate upright* diploid; o, *ouroboros* mutant; oo, ooo, oooo, diploid, triploid and tetraploid homozygous *ouroboros* mutants; (U), presence of one U sex chromosome; (V), presence of one V sex chromosome. The genomes of strains indicated in red have been sequenced.

(B) Graph showing the read depth measured when Illumina reads generated from the female genome (strain Ec597) were mapped onto the male (strain Ec32) SDR scaffold sctg439and285.

Figure S2. Analysis of *Ectocarpus* sp. SDR, PAR and autosomal genes. Related to Figure 2.

(A) Over-representation of DNA corresponding to different transposable element classes in sex chromosome domains compared to autosomes. Histogram showing percent of transposable element DNA per kilobase for six different classes of transposable element expressed as a ratio with respect to the value calculated for the autosomes. The data used were the same as for Figure 2A: cumulative length of DNA corresponding to each class of transposable element in each genomic compartment divided by the length (excluding Ns) of the compartment. Values greater than one indicate an increased abundance of the corresponding class of transposon in the sex chromosome domain compared to the autosomes.

(B) Correlation between the expression of male and female gametologues (FPKM). Male and female FPKM data were log transformed, and a linear regression is shown. The expression levels of male and female gametologue pairs were strongly correlated (Pearson = 0.78, $P<0.001$).

(C) Codon usage in *Ectocarpus*. Three sets of optimal codons were obtained from a multivariate analysis: a conservative set of 10 optimal codons (red), a permissive set of 27 optimal codons (blue), and an intermediate set of 14 optimal codons (green), see Supplementary notes for details. Correlation between each gene's Frequency of Optimal Codons (Fop) and its log10-transformed expression (in FPKM) in mature male gametophytes. Spearman's rho values for each set of optimal codons and their corresponding p-values are indicated on the right. The optimal codons are indicated.

(D) Median Fop in coding regions of autosomes, PAR, male and female SDR haplotypes for each set of optimal codons. Error bars indicate 95% confidence intervals around medians.

(E) Percentages of TE sequence in intronic and intergenic regions of *Ectocarpus* sp. autosomes and of the male SDR haplotype (bootstrap r=1000).

Figure S3. Exon-by-exon analysis of synonymous site substitutions between gametologues. Related to Figure 1, Figure 2 and Figure 3.

Male gametologues are shown in blue, female gametologues in pink. Numbers in plain type indicate synonymous site substitution (dS) values between exons; numbers in italics indicate the standard error of the mean (SEM). dS values in green indicate possible gene conversion events.

Figure S4. Expression analysis of SDR genes. Related to Figure 4.

(A) HMG domain alignment. Multiple alignment of the HMG domains of all the HMG domain proteins encoded by the *Ectocarpus* sp. genome together with HMG domains from human and fungal proteins. The human sequences are Sex-determining region Y (HsSRY, CAA37790.1), Sex-determining region Y-box2 (HsSOX2, CAA83435.1) and Nucleolar transcription factor 1 (HsNTF1, 1608205A). The fungal sequences are *Phycomyces blakesleeanus* Sex+ (PbSex+, ABX27912.1) and Sex- (PbSex-, ABX27909.1), and *Schizosaccharomcyes pombe* Ste11 (SpSTE11, CAA77507.1). The *Ectocarpus* sp. Esi0068_0016 gene is located in the SDR. The *Ectocarpus* sp. proteins Esi0032_0053 and Esi0177_0010 are predicted to have two homeodomains (a and b).

(B) Ratio of the abundance of female SDR gene transcripts in diploid gametophytes (genotypically UV and phenotypically male) compared with mature haploid female gametophytes (genotypically U and phenotypically female). Transcript abundance was measured by quantitative RT-PCR.

(C). Ratio of the abundance of male SDR gene transcripts in diploid gametophytes (genotypically UV and phenotypically male) compared with mature haploid male gametophytes (genotypically V and phenotypically male). Transcript abundance was measured by quantitative RT-PCR.

Figure S5. Estimation of the age of the *Ectocarpus* sp. SDR Related to Figure 5.

Box and whisker plot showing calculated dS values for *Ectocarpus*-Stramenopile pairs plotted against published divergence times (from [S1, S2]). Blue boxes show dS values calculated using codeml, red boxes show dS values calculated using yn00. The species used for the pairwise dS analysis were the following: *Laminaria digitata*, *Saccharina japonica*, *Saccharina latissima* and *Lessonia nigrescens* (98 My), *Fucus distichus*, *Fucus serratus*, *Fucus vesiculosus* and *Sargassum binderi* (114.8 My), *Dictyota dichotoma* (182 My), *Nannochloropsis* (400 My), *Phaeodactylum* and *Thalassiossira* (500 My). The respective linear regressions through the origin are shown as solid lines, with 99% confidence intervals drawn in dotted lines. The green dashed lines show the lowest and highest 99% confidence ages from Codeml and yn00, representing an estimate of the age of *Ectocarpus* sp. SDR at around 95-215 My. The ages where the codeml and yn00 estimates overlap are approximately 170-190 My. Please note that for clarity the y-axis is limited to 15 (two data points are not represented).

SUPPLEMENTAL TABLES

Table S1. *Ectocarpus* strains, sequencing and mapping of the SDRs.

Table S1A. Algal strains used in this study.

Strain	**Details**	**Use in this study**	**Sex**	**Reference**
Ec32	Male genome sequenced strain	Male reference strain (genome sequence, qPCR)	male	[S3]
Ec597	Female genome sequenced strain	Female reference strain (genome sequence, qPCR)	female	this study
Ec603	Male inbred line (8 generations)	RNA-seq and sexual dimorphism analyses	male	this study
Ec602	Female inbred line (8 generations)	RNA-seq and sexual dimorphism analyses	female	this study
Ec87	Sister of Ec32	RT-QPCR	female	this study
Progeny of Ec569	60 strains used for the genetic map	Genetic mapping of the SDR	n/a	[S4]
Progeny of Ec702	2000 progeny of the heterozygous strain Ec702	Recombination analysis	n/a	this study
Ec568	Outcrossing line	Comparative genome hybridisation	female	[S4]
Ec 581-591	Diploid homozygous *oro* mutant gametophytes	Dominance of male haplotype over female	Genotypically UV, phenotypically male	[S5]
Ec761-767	Triploid homozygous *oro* mutant gametophyte (oro,oro,oro)	Dominance of male haplotype over female	Genotypically UUV; phenotypically male	this study
Ec tetraZ4	Tetraploid homozygous *oro* mutant gametophyte (oro,oro,oro,oro)	Dominance of male haplotype over female	Genotypically UUUV; phenotypically male	this study
RB1	Field collected gametophyte	1a strain (Esil 1a_male) used in Figure 5.	Male	this study
DA1	Field collected gametophyte	1a strain (Esil 1a_female) used in Figure 5.	Female	this study
Ec164	Laboratory-produced gametophyte	5b strain (Efas_male) strain used in Figure 5.	Male	this study
Ec184	Laboratory-produced gametophyte	5b strain (Efas_female) strain used in Figure 5.	Female	this study

Sfir_J11 male	Laboratory-produced gametophyte	Male *Sphaerotrichia firma* (E. Gepp) Zinova strain used in Figure 5.	Male	this study
Sfir_O1 female	Laboratory-produced gametophyte	Female *Sphaerotrichia firma* (E. Gepp) Zinova strain used in Figure 5.	Female	this study
Ldig male	Laboratory-produced gametophyte	Male *Laminaria digitata* (Hudson) Lamouroux strain used in Fig. 5	Male	this study
Slom female	Field collected gametophyte	Female *Scytosiphon lomentaria* (Lyngbye) strain used in Fig. 5	Female	this study

Table S1B. PCR-based markers used to map male SDR scaffolds.

N°	Marker	Scaffold	Coordinates on scaffold	Primer 1	Primer 1 sequence	Primer 2	Primer 2 sequence
1	M_68_3	sctg_68	351402	M_68_3F	GTGAGAGAAACAACAGAGCAATACAG	M_68_3R	ATGGAACCGCAGACAACAAGC
2	sctg68_end3	sctg_68	677743	sctg68_end3F	TGGCAACAACACCACCTTTA	sctg68_end3R	AGGACCAGTGAACCATCCTG
3	M_68_4	sctg_68	651823	M_68_4F	AAACACCTCCCAACCAACCAATC	M_68_4R	AACGCAACGAGCAACCTTCC
4	sctg68_5bis	sctg_68	15181	sctg68_5bisF	GGCGCGAAGAAATATATGGA	sctg68_5bisR	CGAGAGTACTGGCCTTTTCG
5	M68_27ex4	Sctg_68	214090	M68_27ex4F	TGCTTGCAACTGGTGACTTC	M68_27ex4R	CGATCATGATGGCAACAACT
6	M68_35ex2	sctg_68	253503	M68_35ex2F	CGTGAACCAGCTTATGGTCA	M68_35ex2R	TCGTTGCAACATCCCAGTAA
7	M_285_1	sctg_285and439	145553	M_285_1F	CAAGCCAACCCACAACTTTT	M_285_1R	GTTGTCATTCGGCGTAGGAT
8	M_439_1	sctg_285and439	95398	M_439_F	CACCACCGTCACGTATGAAG	M_439_R	AGCCCTTGTAGACCCAAGGT

Table S1C. Evidence used to link male scaffolds.

N°	Scaffold 1	Scaffold 1 orientation	Scaffold 2	Scaffold 2 orientation	Scaffolds linked by a miniBAC	Gametologue	Scaffold spanning gene	Scaffold spanning cDNA detected by RT-PCR	RT-PCR primer 1	RT-PCR primer 1 sequence	RT-PCR primer 2	RT-PCR primer 2 sequence
1	sctg_439	antisense	sctg_285	sense	KY0AFIPA75YN02	FeV4scaf04_1	Esi0285_0001	yes	439-285aF	CGATGGCGAAATAAAAGTGG	439-285aR	AGGTTGGAAATTGTGCTTGG

Table S1D. List of strains of known phenotypic sex (determined by crosses with reference strains) that were used to verify sex linkage of male and female SDR scaffolds.

Strain	Species	Lineage	Locality	Phenotypic sex	Genotypic sex	Marker used
Ecsil Nap EA1 f	Esil	1a	Naples, Italy	f	f	68_27ex4, 68_35ex2, 07238, Fe6_15ex5
Ecsil Nap D-A2 f	Esil	1a	Naples, Italy	f	f	68_27ex4, 68_35ex2, 07238, Fe6_15ex5
Ecsil Nap 108 f	Esil	1a	Naples, Italy	f	f	68_27ex4, 68_35ex2, 07238, Fe6_15ex5
Ec597	Esp	1c	San Juan de Marcona, Peru	f	f	68_27ex4, 68_35ex2, 07238
Ec25	Esp	1c	San Juan de Marcona, Peru	f	f	68_27ex4, 68_35ex2, 07238
Ec568	Esp	1c	Arica, Chile	f	f	68_27ex4, 68_35ex2, 07238
Ec87	Esp	1c	San Juan de Marcona, Peru	f	f	68_27ex4, 68_35ex2, 07238
Ec467-U15-2	Esp	1c	San Juan de Marcona, Peru	f	f	68-4, 439_01, scaffold14696
Ec467-U13-5	Esp	1c	San Juan de Marcona, Peru	f	f	68-4, 439_01, scaffold14696
Ec467-U14-2	Esp	1c	San Juan de Marcona, Peru	f	f	68-4 , 439_01, scaffold14696
Ec467-U19-5	Esp	1c	San Juan de Marcona, Peru	f	f	68-4 , 439_01, scaffold14696
EcPH11 85	Ecro	2c	Roscoff, France	f	f	68_27ex4, 68_35ex2, 07238
EcPH11 113	Ecro	2c	Roscoff, France	f	f	68_27ex4, 68_35ex2, 07238
Ec32	Esp	1c	San Juan de Marcona, Peru	m	m	68_27ex4, 68_35ex2, 07238
Ec467-U18-4	Esp	1c	San Juan de Marcona, Peru	m	m	68-4 , 439_01, scaffold14696
Ec467-U9-1	Esp	1c	San Juan de Marcona, Peru	m	m	68-4 , 439_01, scaffold14696
Ec467-U16-5	Esp	1c	San Juan de Marcona, Peru	m	m	68-4 , 439_01, scaffold14696
EcPH11 106	Ecro	2c	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238

EcPH11 112	Ecro	2c	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238
Tam18b	Esp	2d	Tampa, FL, USA	m	m	68_35ex2; 07238
Bft15b	Esp	2d	Beaufort, NC, USA	m	m	68_35ex2, 07238
Tam19c	Esp	2d	Tampa, FL, USA	m	m	68_35ex2; 07238
Tam12b	Esp	2d	Tampa, FL, USA	m	m	68_35ex2;07238
Tam2b	Esp	2d	Tampa, FL, USA	m	m	68_35ex2; 07238
Ecsil Nap R-B1 m	Esil	1a	Naples, Italy	m	m	68_27ex4, 68_35ex2, 07238
Ecsil Na 70 m	Esil	1a	Naples, Italy	m	m	68_27ex4, 68_35ex2, 07238,
Ecsil Na 166 m	Esil	1a	Naples, Italy	m	m	68_27ex4, 68_35ex2, 07238
Ecsil Nap EA2 m	Esil	1a	Naples, Italy	m	m	68_27ex4, 68_35ex2, 07238
PH11-138-1	Esil	1a	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238
PH11-138-6	Esil	1a	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238
PH11-133-1	Esil	1a	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238
PH11-s#2A-38U1-1	Esil	1a	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238
PH11-s#2A-38U1-7	Esil	1a	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238
PH11-s#2A-38U1-3	Esil	1a	Roscoff, France	m	m	68_27ex4, 68_35ex2, 07238

Esil, *Ectocarpus siliculosus*; Ecro, *Ectocarpus crouanorium*; Esp, *Ectocarpus* sp. (no species name attributed); f, female; m, male. Lineages are based on Stache-Crain *et al*. [6].

Table S1E. Assembly statistics for the genome sequence of the female *Ectocarpus* sp. strain Ec597. Data is provided for the intermediate V3 assembly, the final V4 assembly and an assembly carried out with the CLC assembler using only 454 and paired-end Illumina data.

		V3 assembly	V4 assembly	CLC assembly
	Assembler	Velvet	Velvet+ GapCloser	CLC
	Sequence data	454 + paired ends + one mate pair library	454 + paired ends + both mate pair libraries	Paired end data + 454
Contigs (>200 bp)	Number	6 7301	3 3071	98 137
	Avg size (bp)	2373	4648	2 025
	N50 size (bp)	3706	7075	5 052
	Cumulative size (bp)	159 686 990	153 699 260	198 775 585
Scaffolds (>200 bp)	Number	44110	18835	n/a
	Avg size (bp)	5402	12107	n/a
	N50 size (bp)	16323	25114	n/a
	Cumulative size (bp)	23 8264 795	228 041 014	n/a

Table S1F. Evidence used to link female scaffolds.

N°	Scaffold 1	Scaffold 1 orientation	Scaffold 2	Scaffold 2 orientation	Scaffolds linked by the CLC assembly	Scaffold spanning gametologue match	Scaffold spanning gene	Scaffold spanning cDNA detected by RT-PCR	RT-PCR primer 1	RT-PCR primer 1 sequence	RT-PCR primer 2	RT-PCR primer 2 sequence
1	FeV4Scaf10	antisense	FeV4Scaf14	sense	yes	Esi0068_0035	FeV4scaf10_1	yes	FeV4Scaf10-04aF	GCAGTTCTGGAAAACGGAAA	FeV4Scaf10-04aR	GTAGGCGCTTGGTGAAGAGA
2	FeV4Scaf15	antisense	FeV4Scaf6	sense	yes	Esi0068_0003	FeV4scaf15_1	yes	FeV4Scaf15-06aF	AACTGGCGAAGTGGACGA	FeV4Scaf15-06aR	ATACCTGTCCGGCTCCTTG
3	FeV4Scaf7	sense	FeV4Scaf26	sense		Esi0068_0025	FeV4scaf07_6	yes	FeV4Scaf07-26aF	GCGACATACGCACTGTTTTG	FeV4Scaf07-26aR	GGGCCTATCCTCCTCAATGT
4	FeV4Scaf23	sense	FeV4Scaf24	sense			FeV4scaf23_1	yes	FeV4Scaf23-24fF	CTCCCAACGGTCTTGTTCTC	FeV4Scaf23-24fR	GAAAAGTGCATGGTCCCAAC
5	FeV4Scaf57	sense	FeV4Scaf28	sense		Esi0285_0010	FeV4scaf28_2	yes	FeV4Scaf28-57fF	CAATCGGAGATCGTGCTGT	FeV4Scaf28-57fR	GCTTCACGTCCACCTCCTC
6	FeV4Scaf8	antisense	FeV4Scaf9	antisense		Esi0285_0020	FeV4scaf08_1	yes	FeV4Scaf09-08aF	GCCTGGAAGGAGTGAAGGA	FeV4Scaf09-08aR	ACCTGGGATCAATGTTGTCTG
7	FeV4Scaf30	antisense	FeV4Scaf18	sense			FeV4scaf30_1	yes	FeV4Scaf18-30fF	CTGAGTGGAACTGACGTGGA	FeV4Scaf18-30fR	CGACCCTCCTTGATACGTTG
8	FeV4Scaf20	sense	FeV4Scaf4	sense			FeV4scaf04_1	yes	FeV4Scaf20-04aF	CCTGGTTGTTGTGCTTGGA	FeV4Scaf20-04aR	TGCCATCTTTTCACGTCTGT

Table S1G. List of female SDR scaffolds showing their sizes in base pairs and the number of individuals used to map each scaffold to the sex locus.

Scaffold name	length	Carries a gametologue?	Annotated gene on scaffold?	Size of mapping population
FeV4Scaf01	98133	yes	yes	17
FeV4Scaf02	17680	yes	yes	25
FeV4Scaf03	51021	yes	yes	17
FeV4Scaf04and20	78753	yes	yes	23
FeV4Scaf06and15	45363	yes	yes	15
FeV4Scaf07and26	57992	yes	yes	15
FeV4Scaf08and09	41097	yes	yes	8
FeV4Scaf10	14559	yes	yes	8
FeV4Scaf11	25512	no	no	28
FeV4Scaf12	6590	no	no	57
FeV4Scaf17	16002	no	no	55
FeV4Scaf18and30	33261	yes	yes	28
FeV4Scaf19	19637	no	yes	57
FeV4Scaf22	16900	yes but fragment	yes	28
FeV4Scaf24and23	37411	no	yes	55
FeV4Scaf25	63333	yes	yes	18
FeV4Scaf27	52265	no	yes	55
FeV4Scaf28and57	66706	yes	yes	16
FeV4Scaf29	11350	no	no	55
FeV4Scaf31	11604	no	no	55
FeV4Scaf34	2875	no	no	55
FeV4Scaf35	3359	no	yes	18
FeV4Scaf50	3460	no	no	55
FeV4Scaf52	24537	no	yes	55
FeV4Scaf53	18312	no	no	55
FeV4Scaf58	42520	no	yes	55
FeV4scaffold9803	18644	no	no	8
FeV4scaffold8844	2638	no	no	8
FeV4scaffold6686	5237	no	no	8
FeV4scaffold6615	2092	no	no	8
FeV4scaffold6486	2793	no	no	8
FeV4scaffold4676	4105	no	no	8
FeV4scaffold1991	5977	no	no	8
FeV4scaffold17873	16244	no	no	8
FeV4scaffold14543	2896	no	no	8
FeV4scaffold13383	2762	no	no	8
FeV4scaffold12431	1794	no	no	8
FeV4scaffold2006	2715	no	no	8
FeV4scaffold14762	1331	no	no	8
Total	929460			

Table S1H. PCR-based markers used to map female SDR scaffolds.

N°	Marker	Scaffold	Primer 1	Primer 1 sequence	Primer 2	Primer 2 sequence
1	scaffold15594	FeV4scaf01	scaffold15594F	AGACGCGAAGAACGAACACT	scaffold15594R	CCGCGATTTTGTGCTCGTAG
2	scaffold15594a	FeV4scaf01	scaffold15594aF	TTCGCTTTGATTGGGCTATG	scaffold15594aR	ACCAAGTTTCTGGGCAAGGT
3	scaffold5692Da	FeV4scaf01	scaffold5692DaF	GCTCGGCTTGATAGGTCATG	scaffold5692DaR	AGGTTATTGGCCTTGGTTGC
4	Ecfemscaf72511	FeV4scaf01	Ecfemsca72511R	GGTCGGAGAGCGTAAGAGGT	Ecfemsca72511L	TAGGGTTGTTTTGCGATGGA
5	Ecfemscaf4078	FeV4scaf01	Ecfemsca4078R	TAGGGTTGTTTTGCGATGGA	Ecfemsca4078L	GGTCGGAGAGCGTAAGAGGT
6	Ecfemscaf70269	FeV4scaf01	Ecfemsca70269R	GTCGTGGTTGTTCGTGTGTG	Ecfemsca70269L	GTGAGCATTTGGCTGGAAGA
7	Ecfemscaf2091	FeV4scaf01	Ecfemsca2091R	ACCAAGTTTCTGGGCAAGGT	Ecfemsca2091L	TTCGCTTTGATTGGGCTATG
8	Ecfemscaf1174	FeV4scaf01	Ecfemsca1174R	ACCAAGTTTCTGGGCAAGGT	Ecfemsca1174L	TTCGCTTTGATTGGGCTATG
9	Ecfemscaf82697	FeV4scaf01	Ecfemsca82697R	GGGTCGTTCTTTCTGTGCTG	Ecfemsca82697L	TTCAGTTTTCATGCCGTTCC
10	06636	FeV4scaf01	06636F1	GACGCAACAGGGAGGCACCAATA	06636R1	TGCGCGTAACAGGGGAAAAACAA
11	scaffold4647	FeV4scaf02	scaffold4647F	GAATCGGGCTCACGAGAGAG	scaffold4647R	ACGAATTGATTAAGCGGCGC
12	scaffold4647a	FeV4scaf02	scaffold4647aF	CAGGTGGGTTGTCATGTGTG	scaffold4647aR	TACCTACGCCCGAATGAATG
13	scaffold14727	FeV4scaf02	scaffold14727F	GGCGCGGTGAAATACGTTAC	scaffold14727R	GAGGATCGGCAAAATCGCAC
14	scaffold14727a	FeV4scaf02	scaffold14727aF	ACGGTAGGGTCGGAATCAAG	scaffold14727aR	TTGCATGTGTGCGAGTCTGT
15	Ecfemscaf61128	FeV4scaf02	Ecfemsca61128R	GGCAGACCACAACAGGGTAG	Ecfemsca61128L	TAAGCAAGGCTCAACCAGGA
16	07238	FeV4scaf02	07238F1	AAGGAACGCAAACCGCCGAAATA	07238R1	CTCCATCCCCAACGTTGTCTCTGTG
17	scaffold18795a	FeV4scaf03	scaffold18795aF	TTGGCGAAACGAAATCAAAG	scaffold18795aR	TGGTGTAATCGTCCCTGCTC
18	Ecfemsca1372	FeV4scaf04and20	Ecfemsca13724R	ATTTCTGGTGAAAGCGCAAA	Ecfemsca13724L	CACGAAGGAGGGGGTAAAAA
19	Ecfemsca7690	FeV4scaf04and20	Ecfemsca7690R	TGCGACGAGAAAGAAGGAAA	Ecfemsca7690L	AATTGAAACCCCGTCCAATC
20	Ecfemsca5503	FeV4scaf04and20	Ecfemsca5503R	TGCGACGAGAAAGAAGGAAA	Ecfemsca5503L	AATTGAAACCCCGTCCAATC
21	Ecfemsca30656	FeV4scaf04and20	Ecfemsca30656R	TCTTCCAGACGGTGGAGTTG	Ecfemsca30656L	TGGTTACGGCAGCTTCATTT
22	scaffold14581	FeV4scaf04and20	scaffold14581F	CTGTGATTGTTGCGCACACA	scaffold14581R	CGTGTGAGTGGTTTTGGCTG
23	scaffold14581b	FeV4scaf04and20	scaffold14581bF	CGGTACTCCCTCACCACTCA	scaffold14581bR	GGGCAAAAGAGCAACACAAA
24	scaffold10110a	FeV4scaf06and15	scaffold10110aF	TTTGTTGTTGGACCCCTTTG	scaffold10110aR	CCTTCGTTTTCCTCCTCTGG
25	scaffold4518	FeV4scaf06and15	scaffold4518F	GATCCGTTGTTGGCGTATGC	scaffold4518R	GTGCAACGTGCCTGGAATTT
26	scaffold16822a	FeV4scaf06and15	scaffold16822aF	GCCAACACAGCACTACACGA	scaffold16822aR	GGGAATAAACACGACCAGCA
27	scaffold4518a	FeV4scaf06and15	scaffold4518aF	TTGCGGTACTTGTTGCTGTG	scaffold4518aR	AGCGGGAACTAAGCGGTAGA
28	Ecfemscaf15535	FeV4scaf06and15	Ecfemsca15535R	TTGTTTCATCGGCAAAAACC	Ecfemsca15535L	CAGGAACCCCCACTGTATGA
29	scaffold7067	FeV4scaf07and26	scaffold7067F	GAAACGAGCAACGATCGACG	scaffold7067R	TCCACTTTCACTGACGACGG
30	Ecfemsca109773	FeV4scaf07and26	Ecfemsca109773R	GTGCGATACACCGAACGAAC	Ecfemsca109773L	CTCCAATCCCCCACTCATTT
31	scaffold18979	FeV4scaf07and26	scaffold18979F	AAACGACCGCAAGAAACACG	scaffold18979R	GATTCGGCGCACATTCAGTC
32	Ecfemsca3352	FeV4scaf08and09	Ecfemsca3352R	CTCTGACGAGCCACATCCTG	Ecfemsca3352L	GTTCTGGACAACGGTGGAAC
33	Ecfemsca20136	FeV4scaf08and09	Ecfemsca2013R	CGCACGTAGCTCTTTCGATG	Ecfemsca2013L	AAGGTTGTGCTAGGGGGAGA
34	Ecfemsca23188	FeV4scaf08and09	Ecfemsca23188R	CGCATTCCGATTCTCTCCTC	Ecfemsca23188L	CTGCATTCCTCACTCGTTCC
35	Ecfemsca65052	FeV4scaf10	Ecfemsca5052R	GCCTTCCGTGTGCTAGTCTG	Ecfemsca5052L	GCGTGGGTAGATGCAGTAGG
36	Ecfemscaf8896	FeV4scaf11	Ecfemsca8896R	CGCGACCCCCTATCTACTTC	Ecfemsca8896L	AACGCTTCGGAGACTTCACA
37	Ecfemscaf95028	FeV4scaf11	Ecfemsca95028R	GGTTCGGTCTCTCCTGTTCC	Ecfemsca95028L	CCTCTATTGGCGGACCATCT
38	Ecfemscaf26954	FeV4scaf11	Ecfemsca26954R	CGCGACCCCCTATCTACTTC	Ecfemsca26954L	AACGCTTCGGAGACTTCACA
39	scaffold2938	FeV4scaf12	scaffold2938F	CGTTCGTGACGCAATCGTAC	scaffold2938R	CATCCATCCGACGGAAGAGG
40	scaffold2938a	FeV4scaf12	scaffold2938aF	GCGCTGATTGGAAGTGAAAA	scaffold2938aR	CGCAACAACACAAAGAGCAG
41	scaffold12304	FeV4scaf17	scaffold12304F	CTATCCTTCCCGCCTCGAAC	scaffold12304R	GCGAACCTGCGTTGTCTTTT
42	scaffold12304a	FeV4scaf17	scaffold12304aF	AATTTCAGCTCGCCAAGACA	scaffold12304aR	TCTCCCGTTCGGCTATTTTT
43	scaffold19158	FeV4scaf18and30	scaffold19158F	TCCGACCAAGTCCTCGTTTG	scaffold19158R	CGCCAGCGATTCTAACAACG
44	scaffold13317	FeV4scaf18and30	scaffold13317F	GATGCTCGTTCGTTCGTTCG	scaffold13317R	CCATACCCGCATCCTCAGAC
45	scaffold13317a	FeV4scaf18and30	scaffold13317aF	TTACAACAGCCCACCTCACC	scaffold13317aR	GACTGGCGTACCGAAAACAA
46	scaffold1585a	FeV4scaf18and30	scaffold1585aF	TGTTGCTTGTGCGACTGTTC	scaffold1585aR	GTCTCTTGGTGATCGCTTGC
47	scaffold14014a	FeV4scaf19	scaffold14014aF	TGTAAGCGAAGGGAGCAAGA	scaffold14014aR	AGTCTAAGGGCCGGAAACAG

48	scaffold14775a	FeV4scaf22	scaffold14775aF	GCGGTTGAAAGGAAGAGGAG	scaffold14775aR	AAAAAGGCGAAATGGAGGAA
49	scaffold14775	FeV4scaf22	scaffold14775F	CTCGCTCCCGCTTATGTGAT	scaffold14775R	TGTCGAAACGCTTGCTGTTG
50	scaffold7638	FeV4scaf22	scaffold7638F	GCGAAAACCGACGAAACACA	scaffold7638R	CATTCGGTGTTTCGATCCGC
51	scaffold14696	FeV4scaf24and23	scaffold14696F	GTGCGGCGAAAATAATCCCC	scaffold14696R	TTGCTTTCTCTGTTGCACGC
52	scaffold6658	FeV4scaf24and23	scaffold6658F	AAACGAGCCGAGAGATAGCG	scaffold6658R	CCACGAGCTTTGTGTTGGTG
53	scaffold17544e	FeV4scaf24and23	scaffold17544eF	ACCGCAAAACAAAAGTGGAA	scaffold17544eR	GCATTCAGGAGCAAGGAGTG
54	scaffold6658a	FeV4scaf24and23	scaffold6658aF	TCACCAACACAAAGCTCGTG	scaffold6658aR	CGTGTAACGGCTGCATTTTT
55	scaffold7164	FeV4scaf25	scaffold7164F	TATGACGCGTGGCCATAGAC	scaffold7164R	CTGTTCGTTCTCCCCGTGAA
56	scaffold13849	FeV4scaf25	scaffold13849F	GCGAAGATCGAGATCCGGTT	scaffold13849R	GCGTTCGCAGTAACATCGAC
57	scaffold13849a	FeV4scaf25	scaffold13849aF	AACGACGGTAGCTTGGGTTT	scaffold13849aR	TCGGTCGGTTTGTGATTTTC
58	scaffold16354	FeV4scaf27	scaffold16354F	CGCGCAGACCTAGTCATCAT	scaffold16354R	CGCGCGTCCACTTAAGAATG
59	scaffold10828	FeV4scaf28and57	scaffold10828F	AAGAGGACGATGGCTATGCG	scaffold10828R	GAAGCGTTCATCGGCGTAAC
60	scaffold17248a	FeV4scaf28and57	scaffold17248aF	CACTCCGTGAACTTGAACCA	scaffold17248aR	TCGAAGGAGGAAGCAACAAC
61	scaffold17248b	FeV4scaf28and57	scaffold17248bF	CTTTGTGCTTGGGTTGGATT	scaffold17248bR	ATGTCCGCTCTGTCTTTTCC
62	scaffold16559	FeV4scaf29	scaffold16559F	GCCTTTCGCGATGACATCAC	scaffold16559R	GTGAGACGGCCATTCACTGA
63	scaffold14997	FeV4scaf31	scaffold14997F	CATCGATTCAACCTGCAGCG	scaffold14997R	CCACACTACGGACGATAGCC
64	scaffold7781	FeV4scaf34	scaffold7781F	GCCGGCACATCTACCTAGAC	scaffold7781R	TCGAACCGGTGGTCTTTCTG
65	scaffold4943	FeV4scaf35	scaffold4943F	CGCATACTTTGTCGAGTGCG	scaffold4943R	AATCGGAGCCACATCAACGT
66	scaffold1020	FeV4scaf50	scaffold1020F	TTTTGAAAGCGTTCTGCCCG	scaffold1020R	AAACACGATGGGGCTTTTGC
67	scaffold7381	FeV4scaf52	scaffold7381F	GCGAAAACCGACGAAACACA	scaffold7381R	CATTCGGTGTTTCGATCCGC
68	scaffold8468	FeV4scaf53	scaffold8468F	CTTACTTGCGCATGTCGTCG	scaffold8468R	AGGCCGTGAACCACATTAGG
69	scaffold18586	FeV4scaf58	scaffold18586F	TTCTCGGGACAGAAACCACG	scaffold18586R	GCAGCACGATGCACATACAG
70	scaffold12431	FeV4scaffold12431	scaffold12431F	GAGAGGACAGAGGCACACAC	scaffold12431R	ACTCCGAAGCGTGTGATTGT
71	scaffold13383_1	FeV4scaffold13383	scaffold13383_1F	GAGAATCCCTCCGTTGTGTC	scaffold13383_1R	CCACTGCCTGATGTATGGAA
72	scaffold14543	FeV4scaffold14543	scaffold14543F	GCACGAAATACTGCTGCGAG	scaffold14543R	AGAAGACTCGTTGCAGGCTC
73	scaffold14762	FeV4scaffold14762	scaffold14762F	GCGACACACCACTCACAAAC	scaffold14762R	TACGGTGCCCCATTTGTGAA
74	scaffold17873_1	FeV4scaffold17873	scaffold17873_1F	TCTCAAGTACCGCACACACC	scaffold17873_1R	CGCCATTTGATTTTCATCGT
75	scaffold1991_1	FeV4scaffold1991	scaffold1991_1F	CTACGGAACGCCTGAATACC	scaffold1991_1R	GTAACTTTCAAGGGCGACGA
76	scaffold2006	FeV4scaffold2006	scaffold2006F	TAACTTCGACCACATGCCCC	scaffold2006R	TGTTTGCAATGTCAGCCACG
77	scaffold4676	FeV4scaffold4676	scaffold4676F	TGAAACCTCCACCCTAACGC	scaffold4676R	TGATGCGTCGACTCTTACGG
78	scaffold6486_2	FeV4scaffold6486	scaffold6486_2F	GCAATCGTTTCGTCTTTGCT	scaffold6486_2R	CCGACACCACGCTTTAATCT
79	scaffold6615	FeV4scaffold6615	scaffold6615F	ACCCCTCCATCCCTTACTCC	scaffold6615R	ACGATTTAGGGTTCTGCGCA
80	scaffold6686_1	FeV4scaffold6686	scaffold6686_1F	CAAGGGGTACGAAAAACGAA	scaffold6686_1R	AGGGGTAGGGTCGTCAAAGT
81	scaffold8844	FeV4scaffold8844	scaffold8844F	AAAACACACACGCACCGATG	scaffold8844R	TAGCTACTCGGAGGTGCCTT
82	scaffold9803	FeV4scaffold9803	scaffold9803F	TGGGGCGCCATTCCTTTTAT	scaffold9803R	CGCCGCAGACAATACGAAAG

Table S1I. Ploidy of *Ectocarpus* sp. strains determined by flow cytometry.

***Ectocarpus* strain**	**Average fluorescence intensity (arbitrary units)**	**Number of cells analysed**	**Ploidy**
Ec32	57	3111	n
Ec560	52	1037	n
Ec581	127	868	2n
Ec761	163	1066	3n

Ec32, haploid wild type gametophyte; Ec560, haploid *oro* mutant strain; Ec581, homozygous, diploid *oro* mutant strain; Ec761, triploid *oro* mutant strain.

Table S1J. Intra-haplotype sequence similarity in the male and female SDRs.

	male SDR	female SDR
Length (bp) of sequence showing similarity>97%	22685	29680
% total SDR	2.47%	3.20%

Table S2. Male and female SDR genes.

Gene	Main ORF (bp)	n. copies in male genome (Blastp)[1]	FPKM male	FPKM female	total n. transcripts (Cuffdiff)	Gametologue[1]	Autosomal or PAR homologue[1]	Function	Gene/Pseudogene/ Non-coding or repeat	upregulated during fertility? (RT-QPCR)	Comment
Esi0068_0003	921	10	22.41	0	3	FeV4scaf15_1	Esi0208_0044* (LG19)	GTPase activating protein	Gene	yes	Full-length GTPase coding sequence with conserved domain.
Esi0068_0016	942	5	16.96	0	2	no hit	Esi0159_0039*	High mobility group protein	Gene	yes	Not conserved enough to determine if coding region is complete.
Esi0068_0017	1566	42	9.80	0	5	no hit	Esi0066_0062*	Conserved hypothetical protein	Gene	yes	Not conserved enough to determine if coding region is complete. No domains found.
Esi0068_0025	2181	4	26.18	0	1	FeV4scaf07_6	Esi0162_0007* (LG21)	Nucleotide-diphospho-sugar transferase domain protein	Gene	yes	Not conserved enough to determine if coding region is complete.
Esi0068_0027	1056	1	24.24	0	2	FeV4scaf25_1	no hit	MEMO-like domain protein	Gene	yes	Full-length MEMO protein coding sequence with conserved domain.
Esi0068_0035	795	6	44.49	0.016	1	FeV4scaf10_1	Esi0176_0011* (LG18)	Chloroplast clp protease P	Gene	yes	Full-length clp protease coding sequence with conserved domain.
Esi0068_0039	3033	1	16.74	0	3	FeV4scaf30_1	no hit	Conserved hypothetical protein	Gene	yes	Probably full-length coding region. Esi0068_0039 only weakly matches FeV4scaf30_1.
Esi0068_0050	1113	7	0.348	0	1	no hit	Esi0003_0246* (LG2)	Conserved hypothetical protein	Pseudogene	no	Not conserved enough to determine if coding region is complete. No domains found.
Esi0068_0052	804	19	2.86	0.016	3	no hit	Esi0004_0119* (LG3)	Conserved hypothetical protein	Gene	no	Full-length protein coding sequence. No domains found.
Esi0068_0058	1314	>250	44.36	0	2	FeV4scaf01_4	Esi0023_0133* (LG17)	STE20-like serine/threonine kinase	Gene	-	Full-length STE-20 protein coding sequence with conserved domain.
Esi0068_0064	540	2	5.74	3.11	1	no hit	Esi0242_0044* (LG30 PAR)	Putative thioesterase/thiol ester dehydrase-isomerase	Gene	-	Full-length protein coding sequence.
Esi0068_0067	1110	2	120.99	0	2	FeV4scaf02_1	Esi0205_0043	Homoaconitate hydratase	Gene	yes	Full-length protein coding sequence.

Esi0068_0068**	609	3	0	0	1	FeV4scaf25_3	Esi0537_0005*	Conserved hypothetical protein	Transposon remnant	-	Similar to Topoisomerase III. Putative transposon.
Esi0068_0071	1110	2	25.83	0	2	no hit	Esi0099_0001* (LG17)	Ubiquitin C-terminal hydrolase	Gene	yes	Full-length protein coding sequence.
Esi0068_0079	2124	88	10.53	0	4	FeV4scaf03_1	Esi0017_0146* (LG24)	Casein kinase	Gene	yes	Coding region is twice as long as casein kinase sequences in the databases.
Esi0439_0009	288	4	5.75	0	3	no hit	Esi0119_0058* (LG23)	Conserved hypothetical protein	Gene	-	Probably full-length coding region. No domains found.
Esi0439_0005	198	2	0.40	0	1	no hit	Esi0026_0059* (LG12)	Related to HSP70	Pseudogene	-	Matches HSP70s but much shorter. Out of frame ORF with protein matches in the Ec32 genome at the 3rd exon.
Esi0285_0001	1299	128	5.50	0	1	FeV4scaf04_1	Esi0024_0131* (LG15)	LRR and regulator of G protein signaling domain protein	Gene	-	Spans sctg_285 and sctg_439. Not conserved enough to determine if coding region is complete. No domains found.
Esi0285_0010	855	29	14.59	0	2	FeV4scaf28_2	Esi0119_0035* (LG23)	Protein phosphatase 2C	Gene	-	Full-length protein coding sequence.
Esi0285_0020	486	20	25.81	0	3	FeV4scaf08_1	Esi0409_0009	RING-type Zinc finger domain	Gene	no	Probably full-length coding region.
Esi0285_0022	1842	7	0	0	1	no hit	Esi0119_0021* (LG23)	Conserved hypothetical protein	Pseudogene	-	Not conserved enough to determine if coding region is complete. No domains found.
FeV4scaf15_1	2547	11	0	44.03	2	Esi0068_0003	see gametologue above	GTPase activating protein	Gene	no	Large (500 aa) N-terminal extension compared to database matches.
FeV4scaf07_6	2247	4	0	20.97	1	Esi0068_0025	see gametologue above	nucleotide-diphospho-sugar transferase domain protein	Gene	no	Not conserved enough to determine if coding region is complete.
FeV4scaf25_1	735	1	0	5.088	1	Esi0068_0027	see gametologue above	MEMO-like domain protein	Gene	no	Full-length MEMO protein coding sequence with conserved domain.
FeV4scaf10_1	795	6	0	179.84	2	Esi0068_0035	see gametologue above	Chloroplast clp protease P	Gene	no	Full-length clp protease coding sequence with conserved domain.
FeV4scaf30_1	2730	0	0	4.133	2	Esi0068_0039	see gametologue above	Conserved hypothetical protein	Gene	-	Identical copy of second exon on V4scaffold13516. Not conserved enough to determine if coding region is complete. No domains

											found. FeV4scaf30_1 only weakly matches Esi0068_0039.
FeV4scaf01_4	1263	>250	0	45.51	2	Esi0068_0058	see gametologue above	STE20 protein kinase	Gene	no	Full-length STE-20 protein coding sequence with conserved domain.
FeV4scaf02_1	1219	2	0	144.25	2	Esi0068_0067	see gametologue above	Homoaconitate hydratase	Gene	-	Full-length protein coding sequence.
FeV4scaf25_3**	672	2	-	-	1	Esi0068_0068	see gametologue above	Conserved hypothetical protein	Transposon remnant	no	Protein is similar size to male gametologue. No domains found.
FeV4scaf03_1	1329	>50	0	2.35	3	Esi0068_0079	see gametologue above	Casein kinase	Pseudogene	no	Truncated casein kinase gene (end corresponds to FeV4scaf22_1).
FeV4scaf22_1	195	1	0	2.58	3	no hit	no hit	Casein kinase fragment	Pseudogene	-	End fragment of a casein kinase gene. Weak similarity with Esi0068_0079.
FeV4scaf04_1	1113	>50	0	5.06	4	Esi0285_0001	see gametologue above	LRR protein	Gene	-	Not conserved enough to determine if coding region is complete.
FeV4scaf28_2	850	25	0	19.30	3	Esi0285_0010	see gametologue above	Protein phosphatase 2C	Pseudogene	no	Probably full-length coding region.
FeV4scaf08_1	204	5	0	8.42	2	Esi0285_0020	see gametologue above	Fragment of RING-type Zinc finger gene	Pseudogene	-	Shorter than the male gametologue.
FeV4scaf01_2	600	1	-	-	1	no hit	Esi0347_0020*	Putative acetyl-coenzyme A transporter	Pseudogene	no	Protein is truncated compared with acetyl-coenzyme A transporters in Genbank.
FeV4scaf03_2	516	6	0.091	0.72	1	no hit	Esi0408_0016*	Conserved hypothetical protein	Gene	-	Not conserved enough to determine if coding region is complete. No domains found.
FeV4scaf04_2	423	3	-	-	1	no hit	Esi0096_0027* (LG30 PAR)	Histidine triad protein	Gene	-	Not conserved enough to determine if coding region is complete. Histidine triad domain (IPR).
FeV4scaf19_1	159	0	0	2.92	1	no hit	Esi0294_0015*	Hypothetical protein	Gene	-	Not conserved enough to determine if coding region is complete.
FeV4scaf24_2	2271	5	0	8.94	2	no hit	Esi0011_0048* (LG9)	Patched domain protein	Gene	no	Not conserved enough to determine if coding region is complete.
FeV4scaf25_2	1053	4	-	-	1	no hit	Esi0079_0093* (LG27)	MULE transposase domain protein	Transposon remnant	-	Contains MULE transposase domain. Protein is truncated

											compared to Esi0079_0093.
FeV4scaf27_1	395	0	0	2.64	1	no hit	no hit	Hypothetical protein	Gene	-	Not conserved enough to determine if coding region is complete.
FeV4scaf28_1	453	1	-	-	1	no hit	Esi0313_0029* (LG30 PAR)	Conserved hypothetical protein	Pseudogene	-	Based on comparison with Esi031-0028, the other half of this gene is on V4.scaffold10906 (equivalent to sctg_313 in the male).
FeV4scaf35_1	663	11	0	1.70	1	no hit	Esi0142_0041* (LG12)	Heat shock protein 70	Pseudogene	no	Incomplete gene at start of scaffold, also appears to be truncated at the C-terminal end.
FeV4scaf52_1	396	7	-	-	1	no hit	Esi0194_0022* (LG19)	Conserved hypothetical protein	Gene	-	Incomplete gene at start of scaffold.
FeV4scaf58_1	762	4	0	3.29	1	no hit	Esi0034_0115* (LG25)	Conserved hypothetical protein	Gene	-	Not conserved enough to determine if coding region is complete.

[1]blastp of protein against male predicted proteome, cutoff $E10^{-4}$. LG, linkage group; *not sex linked; **The Esi0068_0068/FeV4scaf25_3 pair of genes were predicted to correspond to transposon remnants and were not, therefore, counted among the gametologues, LG, linkage group.

Table S3. Primers used to amplify the sequences from an SDR gene and an autosomal region that were employed to construct the phylogenetic trees shown in Figure 5.

Number	Gene	Primer name	Forward sequence	Reverse sequence
1	Esi0068_0003	ExEsi68_3ex5	TCCAGTTTGTGATGGACTCG	TGAATAATGCCAGACACACTCTG
2	FeV4scaf15_1	ExFe15_1ex6	CGTGGTGGACTCATTGACTG	GTGCCAGACATACCCTGTAGAAC
3	5.8S3F	LSU410R	CGACGGATGTCTTGGCTC	TCCTTCGCTTCCCTTTCAG

Table S4. Expression analysis of SDR genes.

Table S4A. Complete list of HMG-domain proteins encoded by the *Ectocarpus* sp. genome.

Gene	Functional Description	IPR domains	Pfam domains	FPKM male	FPKM female	Differential expression RNA-seq?	Best blastn hit in female Ec597 V4 assembly (scaffold/score/E value)
Esi0031_0053	High mobility group protein	High mobility group, superfamily	HMG (high mobility group) box x2	117.2	86.29	no	scaffold8238/46/0.002
Esi0042_0039	High mobility group and SAP domain protein	DNA-binding SAP; High mobility group, superfamily	SAP domain; HMG (high mobility group) box	48.17	37.82	no	scaffold17135 /266/e-69
Esi0048_0086	High mobility group and histone-like transcription factor domain protein	Transcription factor CBF/NF-Y/archaeal histone; High mobility group, superfamily; Histone-fold	HMG (high mobility group) box; Histone-like transcription factor (CBF/NF-Y) and archaeal histone	6.87	9.91	no	scaffold5380/111/8e-23
Esi0063_0083	High mobility group, SNF2 and SLIDE domain protein	SNF2-related; SANT/Myb domain; Helicase, C-terminal; Homeodomain-like; High mobility group, superfamily; Helicase, superfamily 1/2, ATP-binding domain; SLIDE domain	HMG (high mobility group) box; SNF2 family N-terminal domain; Helicase conserved C-terminal domain; SLIDE	17.94	14.03	no	scaffold11914/957/0.0
Esi0068_0016	High mobility group domain protein	High mobility group, superfamily	HMG (high mobility group) box	11.51	0	n.a	_
Esi0159_0039	High mobility group domain protein	High mobility group, superfamily	HMG (high mobility group) box	184.97	154.83	no	scaffold8835/448/e-125
Esi0159_0081	High mobility group domain protein	Structure-specific recognition protein; High mobility group, superfamily; Domain of unknown function DUF1747; SSRP1 domain	Structure-specific recognition protein (SSRP1); Histone chaperone Rttp106-like ; HMG (high mobility group) box	40.90	47.35	no	scaffold3529/490/e-136
Esi0177_0010	High mobility group domain protein	High mobility group, superfamily	HMG (high mobility group) box x2	0.68	0.785	no	scaffold14575/583/e-165
Esi0224_0005	High mobility group and histone-like transcription factor domain protein	Transcription factor CBF/NF-Y/archaeal histone; High mobility group, superfamily; Histone-fold;	HMG (high mobility group) box; Histone-like transcription factor (CBF/NF-Y) and archaeal histone	19.52	18.24	no	scaffold3986/472/e-131
Esi0228_0021	High mobility group and PHD zinc finger domain protein	Zinc finger, RING-type; Zinc finger, PHD-type; High mobility group, superfamily; Zinc finger, FYVE/PHD-type; Zinc finger, RING/FYVE/PHD-type; Zinc finger, PHD-type, conserved site;	Domain of unknown function (DUF1898); PHD-finger	6.91297	3.78089	no	scaffold8714/163/7e-38

		Zinc finger, PHD-finger;					
Esi0276_0022	High mobility group domain protein	High mobility group, superfamily	High mobility group, superfamily	14.6604	12.2964	no	scaffold14232/2139/0.0
Esi0289_0033	High mobility group, SNF2 and SLIDE domain protein	SNF2-related; SANT/Myb domain; Helicase, C-terminal; Homeodomain-like; High mobility group, superfamily; Helicase, superfamily 1/2, ATP-binding domain; ATPase, nucleosome remodelling ISWI, HAND domain; SLIDE domain	Domain of unknown function (DUF1898) ; SNF2 family N-terminal domain; Helicase conserved C-terminal domain; SLIDE; HMG (high mobility group) box	20.546	17.8343	no	scaffold12975/476/e-132
Esi0446_0014	High mobility group domain protein	High mobility group, superfamily	High mobility group, superfamily	8.65888	4.57929	no	scaffold5638/424/e-117

Table S4B. Statistics for the mapping of male and female mature gametophyte RNA-seq reads onto the "hybrid" genome (corresponding to all male Ec32 scaffolds plus female Ec597 SDR scaffolds). The RNA-seq reads were trimmed and filtered as described in the supplemental experimental procedures before mapping.

	Male replicate 1	**Male replicate 2**	**Female replicate 1**	**Female replicate 2**
Total reads	22 428 865	23 642 187	18 668 732	28 667 939
% mapped reads	85,35	86,35	88,70	89,76

Table S4C. Differential gene expression in male and female mature gametophytes, obtained by RNA-seq analysis.

Sex	**Total number of expressed genes**	**Number of genes showing statistically significant sex-biased expression (DESeq)**	**% sex-biased genes**
Male	14192	354	2.5%
Female	14239	234	1.6%

Table S4D. List of primer pairs used for the quantitative RT-PCR analysis of male and female SDR gene expression.

Gene ID	Primer 1	Primer 1 sequence	Primer 2	Primer 2 sequence	Amplicon size	Analysis
Esi0068_0003	Esi0068_0003F	GCGATGATGGTTGGTATGGT	Esi0068_0003R	CATACGTTGGCTCGTGTGTT	88	Male RT-QPCR
Esi0068_0016	Esi0068_0016F	CTCCCGGAAACAAACAATGAA	Esi0068_0016R	GTCTGACCGCGCTTGATAAC	75	Male RT-QPCR
Esi0068_0017	Esi0068_0017F	CGTTCAACCAGGAAGGACA	Esi0068_0017R	CGTCCGAAGCTCTGCACTAT	162	Male RT-QPCR
Esi0068_0025	Esi0068_0025F	GTCCGTATGAATGGCTGGAT	Esi0068_0025R	TTCCTTCGTGTATCGCTTGTT	128	Male RT-QPCR
Esi0068_0027	Esi0068_0027F	CTCGGACTCTGCCTCGAC	Esi0068_0027R	CAGCAGCACACACCAACTTC	216	Male RT-QPCR
Esi0068_0035	Esi0068_0035F	ATTGCTGTAGGCCACCAACT	Esi0068_0035R	GTTGCGTCGTGCATGTATTC	152	Male RT-QPCR
Esi0068_0039	Esi0068_0039F	AGTCAGGTCGACGCACAAG	Esi0068_0039R	GCTCCCAACAGAGGACACC	80	Male RT-QPCR
Esi0068_0050	Esi0068_0050F	CTACTGCCTCCACTACGCTTC	Esi0068_0050R	CTGCTCCAACATCCTCCATT	203	Male RT-QPCR
Esi0068_0052	Esi0068_0052F	GCGGACGTGTGTATTGTGTT	Esi0068_0052R	TCCTTGCTCGATAGGCTCTG	193	Male RT-QPCR
Esi0068_0067	Esi0068_0067F	AAATGATAGGGTACTGGTGGAGAA	Esi0068_0067R	ATACATTCACAGAGGTCAACACG	105	Male RT-QPCR
Esi0068_0071	Esi0068_0071F	AGTACCGTGGAGTTGTGAAGC	Esi0068_0071R	CCTGTCTTATGACGCACTCG	103	Male RT-QPCR
Esi0068_0079	Esi0068_0079F	TCGCGGATGCCAGACTAT	Esi0068_0079R	GTGCCGGTGAGAGACCTTC	111	Male RT-QPCR
Esi0285_0020	Esi0285_0020F	TCAGGCAGCAAGACTGAGG	Esi0285_0020R	CTGAAGTCCGAACAATGAAGG	172	Male RT-QPCR
FeV4Scaf01_2	FeV4Scaf01_2F	GCGTGATGGATGAGTGGAC	FeV4ScaR01_2R	TGCGAAGAAAGTATCGCTTG	187	Female RT-QPCR
FeV4Scaf01_4	FeV4Scaf01_4F	ATTTCGGCTTGGGTGTTG	FeV4ScaR01_4R	GTATGCCTCGCAGTTGGAAG	156	Female RT-QPCR
FeV4Scaf03_1	FeV4Scaf03_1F	TATACGGCTCAAGGCCACTC	FeV4ScaR03_1R	GCTTCTACCGCTCCAACATC	116	Female RT-QPCR
FeV4Scaf07_6	FeV4Scaf07_6F	GAGAGCGCCTGTTGTATTCG	FeV4ScaR07_6R	TTGATCCCATGAACGAACG	145	Female RT-QPCR
FeV4Scaf10_1	FeV4Scaf10_1F	GAAGGAAAGGAGCAAATGG	FeV4ScaR10_1R	CGTCGTTCGCCAGATAAAG	128	Female RT-QPCR
FeV4Scaf15_1	FeV4Scaf15_1F	TCGTAGTGCTGACGGAAGAG	FeV4ScaR15_1R	GGAAGAGATGCGCTAACACC	138	Female RT-QPCR
FeV4Scaf24_2	FeV4Scaf24_2F	AATACTGCGGGTTTCATGGTAG	FeV4ScaR24_2R	GGTAGTTTCCGTTGCTCATCC	178	Female RT-QPCR
FeV4Scaf25_1	FeV4Scaf25_1F	CCGAAAAAGTGGGAAAGAGG	FeV4ScaR25_1R	GGGAGGAGAACTGAACAATCC	107	Female RT-QPCR
FeV4Scaf25_2	FeV4Scaf25_2F	ATGGGATGGGCAGTGTTC	FeV4ScaR25_2R	GTAACGCTGAACCGCAAGTC	181	Female RT-QPCR
FeV4Scaf25_3	FeV4Scaf25_3F	ATCCCCGGTTTGAAGAGAG	FeV4ScaR25_3R	TCGTTACAGCCGTCATATCG	136	Female RT-QPCR
FeV4Scaf28_2	FeV4Scaf28_2F	CAGTTGCCCTATCCGATGAT	FeV4ScaR28_2R	CAGGCCGATCCTAGTCATCT	146	Female RT-QPCR
FeV4Scaf35_1	FeV4Scaf35_1F	ACGGAACGCACATGAACC	FeV4ScaR35_1R	TATGGCGGTGAACTGATCG	164	Female RT-QPCR

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Ectocarpus strains and culture

Table S1A lists the strains used in this study. See also Figure S1A showing the pedigree of the strains used. The female strain used for genome sequencing, Ec597, was derived by crossing the *ouroboros* mutant Ec494 [S5] with a female strain Ec419 (itself derived from a cross between a female strain Ec25 and the *immediate upright* mutant Ec137). Ec25, Ec137 and the male genome sequenced strain Ec32 [S3] are all meiotic offspring of a field sporophyte, Ec17, collected in 1988 in San Juan de Marcona, Peru [S7]. Ec494 is a UV-mutagenised descendant of Ec32. Two near-isogenic male and female inbred lines Ec601 and Ec602 were derived by repeated crossing of male and female progeny of Ec25 and Ec137 for eight generations. Ec569 was derived by crossing the male genome sequenced strain Ec32 with a female outcrossing line Ec568 from Arica in northern Chile [S4]. Ec702 was derived by crossing the *ouroboros* mutant Ec494 with Ec568. *Ectocarpus* strains were cultured as described [S8].

Generation of RNA-seq transcriptome data for male and female gametophytes

RNA-seq analysis was carried out as described in the main text. All biological material was checked under both a binocular microscope and at higher magnification to confirm fertility (presence of plurilocular gametangia) prior to total RNA extraction. RNA-seq reads (100 bp read length, between 21 and 32 million reads per replicate) were trimmed and filtered with the FASTX toolkit (v0.0.13) using a quality threshold of 25 (base calling) and a minimal size limit of 60 nucleotides. We only retained reads in which more than 75% of nucleotides had a minimal quality threshold of 20.

Two assembly methods were used: *de novo* assembly using Trinity (r2012-01-25) [S9] and reference-based assembly using the male genome sequence (Ec32) and a combination of the TopHat (v2.0.8) [S10, S11] and Cufflinks (v2.1.1) [S11, S12] algorithms.

Each of the four datasets (duplicate male and female samples) was treated separately. Reads were mapped onto the 'hybrid' genome (corresponding to all male Ec32 scaffolds plus female Ec597 SDR scaffolds - see section "Annotation of SDR scaffolds") using TopHat (v2.0.8) [S10, S11] and Cufflinks (v2.1.1) [S11, S12]. The multi-hit parameter of Tophat was 20 and maximum intron size was set at 26,000 bp. Table S4B presents the mapping statistics. Mapped reads were assembled using Cufflinks (v2.1.1.). HTSeq-count [S13] was used to compute FPKM value for each gene.

Based on the Cufflinks analysis and statistical testing using DEseq [S14], only about 4% of the genes expressed during this stage of the life cycle exhibited statistically significant sex-biased expression (Table S4C). The RNA-seq data was also exploited to identify sex-linked genomic scaffolds (see below). Further details about the RNA-seq analysis will be published elsewhere.

Identification and mapping of the male SDR

An *Ectocarpus* gene expression microarray based on EST sequences from the genome-sequenced strain (Ec32) has been used to carry out comparative genome hybridizations for several *Ectocarpus* strains

[S15]. This procedure identified regions of the genome that exhibited significant differences between the Ec32 (male) and Ec568 (female) strains, including two scaffolds that were highly polymorphic along their entire lengths (sctg_68 and sctg_439) and several scaffolds that exhibited polymorphism along only part of their length (sctg_15, sctg_285, sctg_474, sctg_595 and sctg_598).

PCR markers (Table S1C) were designed for each polymorphic scaffold using Primer3 [S16] and *in silico* virtual PCR amplifications were carried out using the e-PCR program [S17] to identify oligonucleotide pairs that were predicted to amplify a single region of the male genome. Sex linkage was initially tested using genomic DNA from four male and four female *Ectocarpus* sp. strains (whose phenotypic sex had been determined by crosses with reference strains). Markers that exhibited sex linkage in this preliminary test were then located on the *Ectocarpus* sp. genetic map by genotyping the 60 individuals of the mapping population [S4]. Markers were added to the genetic map using MAPMAKER [S18]. Marker amplification PCRs were performed using the Promega PCR kit GoTaq® Flexi DNA Polymerase in a total volume of 20 µL containing 2 µL of 1:10 diluted DNA, primers at 100 nM, buffer at 1X, $MgCl_2$ at 2mM, dNTPs at 200 µM and 0.5 units of Taq DNA Polymerase. The thermal profile included an initial denaturation step at 95°C for two minutes followed by 30 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 30 seconds. A final polymerisation step was carried out at 72°C for two minutes. For each PCR, we multiplexed an internal positive control (R26S) [S19] to verify the efficiency of the PCR amplification. Using this approach, sctg_68, sctg_285 and sctg_439 were mapped to the *Ectocarpus* sp. SDR. Scaffolds sctg_15, sctg_474, sctg_595 and sctg_598 were not sex linked.

RNA-seq-based and YGS-based searches for additional male SDR scaffolds

The RNA-seq transcriptome data for male and female gametophytes was used to assess completeness of the male SDR. First, all scaffolds that encoded transcripts with sex-biased expression (FPKM>1 in male samples and FPKM<1 in female samples) were identified. These could have corresponded either to male SDR scaffolds or to scaffolds from other regions of the genome that carried genes with male-biased expression patterns. To eliminate the latter, we examined the context of the genes exhibiting male-biased expression and eliminated cases where these genes were surrounded by non-sex-biased genes. This approach yielded nine new candidate sex locus scaffolds (i.e. in addition to sctg_68, sctg_285 and sctg_439) but comparisons with the female genome revealed that eight of these scaffolds exhibited no significant polymorphism between sexes and genetic mapping of the remaining scaffold showed that it was not sex linked. The RNA-seq-based approach therefore identified no additional male SDR scaffolds.

The Y chromosome Genome Scan (YGS) method developed by Carvalho and Clark [S20] was also adapted to search for additional male SDR scaffolds. This method aims to identify male scaffolds to which kmers generated from female genomic DNA reads map at low density. Low density mapping of female kmers indicates that scaffolds are potentially male-specific sequences and therefore candidate male SDR sequences. Fifteen base pair kmer sequences generated from female strain Ec597 Illumina and 454 sequence reads (see section "Sequencing of a female genome" below) were mapped onto the

1,561 scaffolds of the male Ec32 genome assembly (i.e. all scaffolds longer or equal to 2 kbp) [S3]. One hundred and four scaffolds with a minimum of 70% of unmatched single copy sequence were initially retained as candidate male SDR scaffolds. Thirty-six of these scaffolds were eliminated because the Ec597 sequence reads (Illumina and 454) mapped onto these sequences with a coverage of more than 10x. The remaining 38 candidate male SDR scaffolds were analysed by blasting the complete male scaffold sequences against assemblies of male and female sequence data and manual visualisation of the blast alignments. This manual analysis eliminated all but six of the candidate male scaffolds, which included sctg_68 and sctg_439 that had been identified previously (sctg_285 was not identified by the YGS method because it carries an SDR-flanking non-polymorphic region that is present in both males and females). Two of the four remaining scaffolds contained no unique sequence and it was not possible to design PCR markers to test for sex linkage. PCR markers were generated for the two remaining candidate scaffolds and they were shown to not be sex linked. Therefore, as for the RNA-seq-based approach, the YGS approach did not detect any additional male SDR scaffolds.

Approaches used to improve the assembly of the male SDR haplotype

Although efforts to produce a large insert BAC library for *Ectocarpus* sp. have been unsuccessful, a library of mini-BACs with an insert size of about 15 kbp has been generated [S3]. Using BAC end sequence data, this library was screened *in silico* for BACs that spanned two of the three male SDR scaffolds. One BAC was detected (KY0FIPA75YN02) that matched ends of both sctg_285 and sctg_439, allowing these two scaffolds to be fused into a single scaffold (Table S1C). No BACs were detected that linked sctg_68 to either sctg_285 or sctg_439. The colinearity of sctg_285 and sctg_439 was confirmed independently by amplification of cDNA corresponding to the gene Esi0285_0001, which spans the two scaffolds, using oligonucleotide primers that corresponded to exons at the two adjacent ends of the scaffolds. The scaffold formed by assembling sctg_285 and sctg_439 is referred to as sctg_285and439.

Recombination analysis

Recombination between four sex locus markers corresponding to scaffolds sctg_68 and sctg_285and439 (Table S1B) was analysed using a large segregating family of 2000 meiotic individuals (Figure S1A) derived from a cross between the male line Ec494 [S5] and the female outcrossing line Ec568 [S4]. Between 200 and 1500 ng/µL of DNA was extracted from 100 mg of tissue (fresh weight) from each individual in the population using the Nucleospin® Multi-96 plant kit (Macherey-Nagel) according to the manufacturer's protocol and diluted 1:10. PCR reactions were performed using the Promega PCR kit GoTaq® Flexi DNA Polymerase in a total volume of 20 µL containing 2 µL of 1:10 diluted DNA, oligonucleotides at 100 nM, buffer at 1X, $MgCl_2$ at 2mM, dNTPs at 200 µM and 0.5 units of Taq DNA Polymerase. The thermal profile included an initial denaturation step at 95°C for two minutes followed by 30 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 30 seconds. A final polymerisation step was carried out at 72°C for two minutes. An internal positive control (R26S)

[S19] was used to verify the efficiency of PCR amplification. No recombination was detected between any of the markers located within the male SDR (Figure S2).

Sequencing of a female genome

The genome of the female strain Ec597 (Table S1A, Figure S1) was sequenced using a whole genome shotgun strategy using both Illumina HiSeq 2000 technology and Roche 454 pyrosequencing. One paired-end (PE) library with fragment sizes of about 180 bp, and two mate-pair (MP) libraries with insert sizes of 10 kbp were constructed for the Illumina sequencing. In total 18.1 Gbp data of paired-end reads of 104 bp and 9.9 Gbp of mate-pair reads of 51 bp were generated from these libraries.

Velvet (version 1.1.05) was used to run several assemblies during the sequencing process, including the V3 assembly (which used all the PE reads and reads from one of the MP libraries) and the final V4 assembly with the complete read dataset (Table S1E). The V3 assembly is the raw Velvet output, launched with a kmer value of 45. The V4 assembly was generated using Velvet with a kmer value of 51 follow by a step of gap closing using the tools provided with the SOAP *de novo* assembler (GapCloser). An independent *de novo* assembly was also carried out with the CLC assembler (http://www.clcbio.com/products/clc-assembly-cell) using only the paired end Illumina data.

Identification of scaffolds corresponding to the female haplotype of the SDR

Female SDR scaffolds were identified using three different approaches. First the deduced protein sequences of male SDR genes (all annotated genes on the two male SDR scaffolds sctg_68 and sctg_285and439) were blasted against both the V3 and V4 versions of the female genome assembly to detect scaffolds carrying female alleles (gametologues) of the male SDR genes. Fourteen candidate female SDR scaffolds were identified in the V4 assembly using this approach.

The second approach employed the RNA-seq transcriptome data. The two sets of female transcripts constructed by Trinity using the replicate female RNA-seq datasets were independently compared with the male and female genome assemblies by local alignment using Blast. Transcripts that aligned with the female genome but not with the male genome were retained (E value cut-off: $1.e^{-4}$). The two sets of female-specific transcripts from the two replicates were then clustered and the local alignment with the female genome repeated to generate a list of putative female-specific scaffolds. Ninety-seven candidate female SDR scaffolds were identified in the V4 assembly using this approach.

PCR markers were developed for all candidate female SDR scaffolds using Primer 3 [S16] and *in silico* virtual PCR amplifications were carried out using the e-PCR program [S17] to identify oligonucleotide pairs that were predicted to amplify a single region of the female genome but not to amplify from the male genome (Table S1G). Each marker was then tested on genomic DNA of between eight and 57 individuals (at least 18 individuals if the scaffold did not carry a gametologue) of known sex to determine whether the candidate scaffolds were genetically linked to the sex locus (Table S1H). PCR

reactions were carried out as described above but in a final volume of 10 µl. The presence or absence of single sex-specific bands was resolved by electrophoresis on 1.5% agarose gels.

The third approach used involved adapting the Y chromosome Genome Scan (YGS) method [S20] to search for additional female SDR scaffolds. Fifteen base pair kmer sequences from the Sanger reads that had been generated for the male Ec32 genome assembly [S3] were mapped onto the scaffolds of the V4 assembly of the genome of the female strain Ec597. Only scaffolds longer than 200 bp with a minimum of 85% of unmatched single copy sequence were retained. These scaffolds were screened against the Genbank collection of bacteria sequences using Blastn to remove contaminating bacterial DNA sequences. Illumina sequence reads generated from female genomic DNA were mapped onto the retained scaffolds and scaffolds with coverage of greater than 10x were removed. The remaining scaffolds included 28 that had already been identified by the two approaches described above, plus an additional 260 scaffolds that had not been identified previously. The 260 candidate female SDR scaffolds were further analysed by blasting the complete female scaffold sequences against assemblies of male and female sequence data and manual visualisation of the blast alignments. This manual analysis eliminated all but 66 of the candidate female scaffolds. Scaffold-specific markers were developed for the 66 scaffolds using Primer3 [S16] and *in silico* virtual PCR amplifications using the e-PCR program [S17]. Each marker was then tested on genomic DNA of eight individuals of known sex to determine whether the candidate scaffolds were genetically linked to the sex locus. Thirteen additional female SDR scaffolds were validated by this approach. None of the thirteen scaffolds carried a protein-coding gene.

Approaches used to improve the assembly of the female SDR

Sex-determining regions are extremely difficult to assemble because they exhibit a high density of repeats. Several strategies were used to improve the assembly of the female SDR (Table S1G). These strategies were applied in parallel with the genetic mapping tests (see above) and focused on scaffolds that had been shown to be sex-linked by these tests. Scaffold structure differed significantly between the V3 and V4 assemblies of the female genome and iterative reciprocal Blasts between the two assemblies allowed many of the putative female SDR scaffolds to be manually extended and linked together. Additional evidence for links between scaffolds were obtained by 1) mapping of raw mate-pair sequence data (using Bowtie [S21]), 2) looking for matches to male SDR genes that spanned scaffolds and 3) using scaffold-spanning transcripts predicted by Trinity based on the female RNA-seq data. The CLC assembly (Table S1E) also confirmed several links between scaffolds (Table S1F). When a transcript linking two scaffolds was predicted based either on RNA-seq data or similarity with a male gametologue, the transcript was verified experimentally by reverse transcriptase PCR amplification from cDNA of between four and eight females using oligonucleotides corresponding to the exons at the two ends of the linked scaffolds (Table S1F). RNA was extracted using the Qiagen Mini kit (http://www.qiagen.com) as previously described [S5]. RT-PCR reactions were performed using the RT-PCR OneStep kit (QIAGEN, Courtaboeuf, France) following the manufacturers specifications except that we added 5 ng of template RNA and the final reaction volume was 10 µl rather than 50 µl. Thermal

cycles were performed as follows: A single cycle for reverse transcription at 50 °C for 30 min then an initial denaturing /PCR activation step at 95 °C for 15 min followed immediately by 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 57 °C for 30 seconds, extension at 72 °C, and a single final extension step at 72 °C for 10 minutes. Single bands were resolved by electrophoresis on a 10 cm long 2% agarose gel. The application of these various approaches allowed the total number of female SDR scaffolds to be reduced to 40.

Annotation of SDR scaffolds

Previously annotated scaffolds from the *Ectocarpus* sp. genome project [S3] were considerably improved by integrating transcript information derived from the RNA-seq analysis carried out as part of this study and using comparisons of male and female gametologue gene models. The updated gene models can be accessed at the OrcAE database (http://bioinformatics.psb.ugent.be/orcae/overview/Ectsi, [S22]).

The female SDR scaffolds were annotated *de novo* by running the gene prediction program EuGène [S23], which incorporated the signal prediction program SpliceMachine [S24], using the optimised Markov models and SpliceMachine splice site predictions derived previously for the male genome sequence [3]. Gene prediction incorporated extrinsic information from mapping of the RNA-seq data onto the female-specific scaffolds. Both male and female SDR gene models were manually curated using the raw, mapped RNA-seq data, Cuffdiff and Trinity transcript predictions and comparisons between the male and female haplotypes.

The female SDR scaffolds were added to the complete male genome scaffolds to produce a “hybrid genome”, which served as the basis for further reference-based analyses. Assembly of the mature gametophyte RNA-seq data using this hybrid reference genome as a template was carried out using TopHat (v2.0.3) [S10, S11] and Cufflinks (v2.0.2) [S11, S12]. For TopHat, the maximum value for multihits per read was set at 40. For both TopHat and Cufflinks the maximum intron size was set at 26,000 bp and annotations were used to guide mapping and transcript assembly against the "hybrid genome".

A first reference assembly of the mature gametophyte RNA-seq data, using only the male annotations, was performed with this “hybrid genome”. Results of this first assembly were used to confirm automatic gene predictions and attempt to identify new genes. New genes specific to the female were annotated and an annotation file was created. A second assembly was then carried out using the “hybrid genome” reference with both the male and the new female annotation files to compute the abundance of female genes (FPKM) with more accuracy.

Pseudogenes were identified manually by comparing SDR sequences with genes in the public databases. An additional screen for pseudogenes was carried out by blasting male protein sequences against the genomic sequence of the female SDR and vice versa. All sequences that had been annotated as “gene” or “TE” were excluded from this latter analysis using Maskseq and RepeatMasker respectively. We defined a pseudogene operationally as a fragment of nucleotide sequence that

resembled a protein sequence in the public databases but was truncated due to the presence of stop codons or frameshifts.

Figure 3 provides a schematic overview of the numbers of loci annotated in the male and female SDR haplotypes, together with information about homology relationships between SDR loci and with autosomal genes. The male SDR haplotype contains 17 protein coding genes and three pseudogenes, the female 15 protein-coding genes and seven pseudogenes. Three additional loci (two female and one male) are probably remnants of transposable elements, despite being single copy in the genome, because they share homology with typical transposon proteins such as transposases. The putative transposon remnants were excluded from the analysis of synonymous site divergence (Figure 5C) and the list of gametologue pairs and also from the gene counts and other statistics (Table 1).

Homologous genes present in both the male and female haplotypes of the SDR were considered to be gametologues (i.e. male and female alleles of the same ancestral gene) if they were detected as matches in a reciprocal Blastp search against the SDR scaffolds (E value cutoff: 1 e^{-4}). The same criterion was used to identify homologues of SDR genes located outside the SDR (Table S2).

Identification of transposons and other repeated sequences in the SDR

An *Ectocarpus*-specific TE-library (described in [S3]), compiled with REPET [S25], was used to annotate SDR transposons. TEs were further annotated by running the *de novo* annotation software Repclass [S26] with default parameters. The annotation data from REPET and Repclass were then merged using a custom script. The script retained the REPET annotation when there was a conflict. The autosomes, the PAR and the male and female SDR haplotypes were screened for TEs by running RepeatMasker on each of these genomic compartments using the TE library described above. A custom script was used to parse the RepeatMasker output and to count the total sequence length of each TE category in each genomic compartment. This value was divided by the total length of each compartment (excluding Ns) to calculate the percentage of each region that corresponded to TE sequence.

Intra-haplotype sequence similarity

Sequence similarities within the male and female SDR haplotypes were analysed using a custom Perl code [S27]. This code used BLAST (http://blast.wustl.edu) and a moving window system to compare 5 kbp sequence segments, in steps of 2 kbp, with the rest of the SDR sequence to detect repeated regions within either the male or the female SDR haplotype (i.e. the analysis was carried out separately for the male and female haplotypes). The sliding window Blast analysis was performed on sequences in which the transposable elements had been masked. By default, the threshold for sequence identity was fixed to 97%. When the threshold was reduced to 50%, similar results were obtained.

Global expression of autosomal versus SDR genes

We compared mean transcript abundances in sexually mature, male and female gametophytes, determined by RNA-seq and expressed as the log2 of FPKM mapped, for genes from different genome fractions. The data did not follow a normal distribution, so we performed a non-parametric test (Wilcoxon) which indicated statistical differences between the mean abundances of transcripts of PAR and SDR genes that are haplotype-specific (asterisks in Figure 2C). When all the SDR genes were analyzed, their transcripts were less abundant, on average, than those of autosomal genes, but the difference was not significant. To take into account possible bias due to the small sample sizes for the PAR genes (n=205), SDR genes (n=37) and the SDR genes excluding the gametologues (n=16), we computed 10,000 individual autosomal p-value distributions (Wilcoxon test) with appropriate sized samples of autosomal genes (205, 37 or 16 genes) and compared these resampled p-values with the observed sample data. The percentage of re-sampled values that had a p-value lower than the observed p-value (p-adjusted) was 0.02%, 30% and 0.7% (corresponding to PAR, SDR and SDR without gametologues datasets respectively). The resampling approach therefore confirmed that transcripts of PAR and SDR genes without gametologues were significantly less abundant on average than transcripts of autosomal genes. When all the SDR genes were analyzed, transcripts were less abundant, on average, than those of autosomal genes, but the difference was less pronounced than when the gametologue genes are excluded.

Quantitative reverse transcriptase PCR analysis of SDR gene transcript abundances during the *Ectocarpus* sp. life cycle

The abundance of male and female SDR gene transcripts during the *Ectocarpus* sp. life cycle was assessed by RT-QPCR. Primer pairs were designed to amplify regions of the 3'UTR or the most 3' exon of the gene to be analysed (Table S4D). *In silico* virtual PCR amplifications were carried out using the e-PCR program [S17] and both the male and female genome sequences to check the specificity of oligonucleotide pairs. RT-QPCR analysis was carried out for 13 male SDR genes and 11 female SDR genes (Figures S4). The remaining SDR genes could not be analysed either because they had very small exons, which posed a problem for primer design, or it was not possible to obtain a single amplification product.

For the RT-QPCR analysis, total RNA was extracted using the Plant RNeasy extraction kit (Qiagen, Courtaboeuf, France) from at least three biological replicates for each of four stages of the life cycle: immature gametophyte, mature gametophyte, partheno-sporophyte and diploid heterozygous sporophyte (Figure 1A). The RNA was treated with RNAse-free DNAse-I according to the manufacturer's instructions (Qiagen) to remove any contaminating DNA and stored at -80°C. The concentration and integrity of the RNA was checked using a NanoDrop 2000 spectrophotometer (ThermoScientific) and by agarose gel electrophoresis. A control PCR without reverse transcriptase was performed to ensure absence of contaminating DNA. For each sample, up to 1 µg of RNA was reverse-transcribed to cDNA using oligo-dT and the Superscript II RT kit (Life Technologies, Gaithersburg, MD, USA) according to the manufacturer's instructions and the cDNA was diluted with water to 1.2 ng equivalent RNA.μL^{-1}.

RT-QPCR was carried out using the ABsolute™ QPCR SYBR® Green ROX Mix (ThermoScientific) in a Chromo4™ thermocycler (BioRad Laboratories) and data were analysed with the Opticon monitor 3 software (BioRad Laboratories) or a LightCycler® 480 multiwell plate 384, on a LightCycler® 480 Real-Time PCR System (Roche Diagnostics, Mannheim, Germany), using the LightCycler® 480 SYBR Green Master mix (Roche Diagnostics, Mannheim, Germany).

For each gene, amplification efficiency (always between 80% and 110%) was tested using a dilution series of male or female *Ectocarpus* sp. genomic DNA (15 ng to 0.006 ng), each dilution being tested in duplicate. Using these genomic DNA dilutions, a standard curve was established for each gene, allowing quantification. Amplification specificity was tested with a dissociation curve. The housekeeping gene *ELONGATION FACTOR 1α* (*EF1α*) [S7] was used to normalise transcript abundance values. For sporophyte samples, values for *EF1α* transcript abundance were divided by two to correct for the difference in copy number of autosomal compared to SDR genes. The normalized data correspond to means ±S.E. from three to four independent biological replicates, each of which was calculated from three technical replicates.

To test for the difference in gene expression between the different life cycle stages (immature and mature gametophyte, partheno-sporophyte and sporophyte), a one-way analysis of variance (ANOVA) was performed for the 13 male and 11 female SDR genes, implemented in MINITAB (version 13.2 MiniTab Inc. 1994, State College USA). Data were log-transformed in order to meet the normality and homoscedasticity requirement of ANOVA and multiple comparisons of means were performed using the Fisher method (for Esi0068_0071, Esi0068_0079 and FeV4Scaf35_1 genes as log-transformed data did not meet the homoscedasticity requirement, the Kruskal-Wallis non-parametric test was performed).

Construction of phylogenetic trees for an SDR gene and an autosomal region

Coding sequences from a single exon of two genes from the same gametologue pair (Esi68_0003 (male) and FeV4scaf15_1 (female)) were amplified in three *Ectocarpus* lineages and three related brown algae species *Sphaerotrichia firma* (E. Gepp) Zinova, *Scytosiphon lomentaria* and *Laminaria digitata*. The *Ectocarpus* strains described as lineages probably represent separate species based on sequence divergence of autosomal genes, morphology and on sexual crossing experiments. We use the term lineage since only three *Ectocarpus* species currently have species status [S28]. Two of the recognised species, *Ectocarpus siliculosus* (lineage 1a) and *Ectocarpus fasciculatus* (lineage 5b) were included in the analysis, and we also used the sequenced strain Ec32 that belongs to lineage 1c. The autosomal tree was constructed using sequences amplified from the same samples using ITS2 nuclear DNA primers. The strains and lineages used are described in Table S1A and the primers are described in Table S3. The PCR conditions were as described above in the "Identification and mapping of the male and female SDR haplotypes" section, but with a final volume of 20 µl to allow for DNA sequencing. Single bands were directly sequenced in the forward and reverse direction. Sequences were edited using Codon Code sequence aligner and the evolutionary history was inferred by constructing

phylogenetic trees using the Maximum likelihood approach (PhyML) implemented in MEGA5 [S29]. PhyML trees were tested with 1000 bootstraps under the Tamura-Nei (1993) model assuming a uniform nucleotide substitution rate. The construction of trees was repeated using the neighbour joining method and the same topology was obtained. The PhyML trees are presented in Figure 5.

Synonymous divergence

To estimate synonymous divergence the coding sequences of gametologue pairs were translated to protein sequence, and an alignment was performed in the program Seaview [S30] using the default Muscle parameters. Alignments were further confirmed using Prank [S31] for verification. Those alignments with large regions that were poorly aligned were further analysed using Gblocks [S32] in order to eliminate these sections. The aligned protein sequences were then back-translated to coding sequence and synonymous divergence (dS) between gametologous gene pairs was calculated using Codeml in PAML version 4 [S33]. The resulting values were plotted against the gene coordinates the male V chromosome in order to visualise the spread of the data according to gene position and detect whether their positions were organised in relation to the degree of divergence from their female counterpart. Analyses of the data using Kmeans in the program Rv3.0.1 was able to resolve both K of one and two. A Mann-Whitney U test of the resolution K2 was not significant therefore the null hypothesis that the data are formed from a single population was retained. This result implied the absence of gene strata within the male SDR haplotype.

To search for potential gene conversion events dS values were calculated on an exon-by-exon basis (Figure S3). Potential gene conversion events could then be identified by looking for marked variations in dS along the length of each gene.

Estimating the age of the *Ectocarpus* sp. SDR

The complete nuclear coding sequences of *Phaeodactylum tricornutum*, *Thalassiosira pseudonana* and *Ectocarpus* sp. (autosomes) were obtained from the Hogenom database version 6 [S34], together with coding sequence data from a further 63 Stramenopile species (following those used in Silberfeld *et al*. [S2] and Brown and Sorhannus [S1]), which were downloaded from Genbank (http://www.ncbi.nlm.nih.gov/genbank/). Clusters of potentially orthologous genes were identified using the program Silix [S35] and deduced protein sequences derived from the above sequences. Alignments of the potential orthologues were then created using Muscle and the topology of the resulting sequence clusters were reconstructed using Phyml under a GTR model with an estimated value for gamma and with five different classes [S36]. These clusters were then analysed using the program TPMS [S37], which can identify the sub-clusters with potential orthologues and hence facilitates the elimination of paralogues. Clusters that contained at least one gene from *Ectocarpus* sp. and one gene from each of the two diatom species, *Phaeodactylum* and *Thalassiosira,* were retained for further analysis.

A blastx search of the *Ectocarpus* sp. genes retained in the clusters was performed against RNA-Seq and Sanger EST data from nine Phaeophyceae species and significant matches (E-value $1e^{-10}$) were added to the cluster data. The phylogenies of the gene clusters were further verified manually and those that did not correspond to the species phylogeny were eliminated. In total, 54 clusters of genes from *Ectocarpus*, the two diatoms and 9 phaeophyceae and 1 Eustigmatophyceae species were obtained, making up 183 pairs. The orthology of the *Ectocarpus* sp. and both diatom sequences was further checked for each of the 54 clusters using the information available in the gene family database Hogenom 6, which includes most the fully sequenced eukaryotic genomes. We searched the phylogenetic trees of the 54 *Ectocarpus* sp. genes in Hogenom 6 and checked whether the diatom sequences that we identified as orthologous were actually the most closely related to the *Ectocarpus* sp. sequence in the those trees. Orthology of the *Ectocarpus* sp. and both diatom sequences was confirmed for 27 clusters, partly confirmed or possible for 10, and not confirmed for the remaining clusters. Only the diatom sequences with confirmed or possible orthology were retained for the subsequent steps of the analysis (i.e. 147 sequence pairs).

A pairwise alignment of the *Ectocarpus* sp. genes with all of the identified orthologous genes from each cluster was then carried out using Prank [S31], and alignments were cleaned using Gblocks with highly stringent parameters (maximum number of contiguous non-conserved sites = 2) as we found that less stringent parameters returned less reliable dS values [S32, S33]. The programs Codeml and Yn00 from PAML version 4 [S33] were then run on each gene pair in order to calculate pairwise dS values. Pairs with aberrant dS-yn00 values (i.e. 99) were excluded, as well as pairs with very high dS-ML values (>20) as these high values probably resulted from convergence problems with divergent sequences in the case of codeml, or to the presence of hidden paralogy among the brown algal sequences. A total of 137 pairs were available for further analysis. The resulting dS values were plotted against the divergence times (Figure S5) estimated by Silberfeld *et al*. [S2] and Brown and Sorhannus [S1].

Codon usage analysis

Optimal codons were identified by comparing the codon usage of highly expressed genes (ribosomal genes) with the rest of the genome using the multivariate approach described in Charif *et al*. [S38] (see also http://pbil.univ-lyon1.fr/members/lobry/repro/bioinfo04/). A conservative set of 10 optimal codons was identified, together with two additional, less conservative sets of 14 and 27 optimal codons (Figure S2C and S2D). A custom script was used to estimate the frequency of these optimal codons (Fop) in *Ectocarpus* sp. coding sequences. The Fop values were correlated with RNA-seq expression levels (Figure S2D). We retained the set of 27 optimal codons for the analysis of the SDR genes as it gave the highest Spearman rho values. The Fop values in each compartment (autosomes, PAR and SDR) for the three sets of codons are shown in Figure S2D.

Sex-determination in strains carrying different numbers of U and V chromosomes

Diploid gametophytes carrying both the U and the V chromosome can be constructed artificially using the *ouroboros* (*oro*) mutant, and these strains are phenotypically male [S5]. To determine whether the dominance of the male haplotype was dose dependent, we constructed triploid (UUV) gametophytes, again using *oro* mutant strains (Figure S1). We isolated 10 independent UV diploid strains (Ec581 to Ec591) and seven independent UUV triploid strains (Ec761 to Ec767) using zygote isolation methods described in Coelho *et al*. [S39]. The ploidy of one representative each of the polyploid *Ectocarpus* sp. strains was verified using a FACSort flow cytometer (http://www.bsbiosciences.com; Table S1I). Nuclei were isolated by cutting the filaments with a razor blade and adding nuclei buffer [30 mM MgCl, 120 mM trisodium citrate, 120 mM sorbitol, 55 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES) pH 8, 5 mM EDTA supplemented with 0.1% (v/v) Triton X-100 and 5 mM sodium bisulphite], and DNA content was measured immediately by flow cytometry. Between 600 and 13,200 nuclei were analyzed for each sample. Gametophytes from the male Ec32 strain were used as an internal reference [S40, S41]. The nucleic acid-specific stain SYBR Green I (http://www.invitrogen.com) was used at a final dilution of 1:10,000. All UV and UUV strains were phenotypically male indicating that the dominance of the male haplotype over the female haplotype is not dose dependent.

HMG-domain genes present in the *Ectocarpus* sp. genome

A survey for HMG-domain genes identified a total of 13 genes in the *Ectocarpus* sp. genome including the male-specific gene Esi0068_0016, which is the only HMG-domain gene inside the SDR (Table S4A, Figure S4A). Based on the gametophyte RNA-seq data, none of the 12 autosomal HMG-domain genes exhibited differential expression in male versus female *Ectocarpus*.

SUPPLEMENTAL REFERENCES

S1. Brown, J.W., and Sorhannus, U. (2010). A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. PLoS One *5*.

S2. Silberfeld, T., Leigh, J.W., Verbruggen, H., Cruaud, C., de Reviers, B., and Rousseau, F. (2010). A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the "brown algal crown radiation". Mol Phylogenet Evol *56*, 659-674.

S3. Cock, J.M., Sterck, L., Rouzé, P., Scornet, D., Allen, A.E., Amoutzias, G., Anthouard, V., Artiguenave, F., Aury, J., Badger, J., et al. (2010). The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. Nature *465*, 617-621.

S4. Heesch, S., Cho, G.Y., Peters, A.F., Le Corguillé, G., Falentin, C., Boutet, G., Coëdel, S., Jubin, C., Samson, G., Corre, E., et al. (2010). A sequence-tagged genetic map for the brown alga *Ectocarpus siliculosus* provides large-scale assembly of the genome sequence. New Phytol *188*, 42-51.

S5. Coelho, S.M., Godfroy, O., Arun, A., Le Corguillé, G., Peters, A.F., and Cock, J.M. (2011). *OUROBOROS* is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga *Ectocarpus*. Proc Natl Acad Sci U S A *108*, 11518-11523.

S6. Stache-Crain, B., Müller, D.G., and Goff, L.J. (1997). Molecular systematics of *Ectocarpus* and *Kuckuckia* (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear and plastid-encoded DNA sequences. J Phycol *33*, 152-168.

S7. Peters, A.F., Scornet, D., Ratin, M., Charrier, B., Monnier, A., Merrien, Y., Corre, E., Coelho, S.M., and Cock, J.M. (2008). Life-cycle-generation-specific developmental processes are modified in the *immediate upright* mutant of the brown alga *Ectocarpus siliculosus*. Development *135*, 1503-1512.

S8. Coelho, S.M., Scornet, D., Rousvoal, S., Peters, N.T., Dartevelle, L., Peters, A.F., and Cock, J.M. (2012). How to cultivate *Ectocarpus*. Cold Spring Harb Protoc *2012*, 258-261.

S9. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol *29*, 644-652.

S10. Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

S11. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc *7*, 562-578.

S12. Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol *28*, 511-515.

S13. Anders, S., Pyl, S.T., and Huber, W. (2014). HTSeq — A Python framework to work with high-throughput sequencing data. bioRxiv preprint.

S14. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

S15. Dittami, S.M., Proux, C., Rousvoal, S., Peters, A.F., Cock, J.M., Coppée, J.Y., Boyen, C., and Tonon, T. (2011). Microarray estimation of genomic inter-strain variability in the genus *Ectocarpus* (Phaeophyceae). BMC Mol Biol *12*, 2.

S16. Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol *132*, 365-386.

S17. Schuler, G.D. (1997). Sequence mapping by electronic PCR. Genome Res *7*, 541-550.

S18. Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., Newberg, L.A., and Newburg, L. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics *1*, 174-181.

S19. Le Bail, A., Dittami, S.M., de Franco, P.O., Rousvoal, S., Cock, M.J., Tonon, T., and Charrier, B. (2008). Normalisation genes for expression analyses in the brown alga model *Ectocarpus siliculosus*. BMC Mol Biol *9*, 75.

S20. Carvalho, A.B., and Clark, A.G. (2013). Efficient identification of Y chromosome sequences in the human and Drosophila genomes. Genome Res *23*, 1894-1907.

S21. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

S22. Sterck, L., Billiau, K., Abeel, T., Rouzé, P., and Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. Nat Methods *9*, 1041.

S23. Foissac, S., Gouzy, J.P., Rombauts, S., Mathé, C., Amselem, J., Sterck, L., Van de Peer, Y., Rouzé, P., and Schiex, T. (2008). Genome Annotation in Plants and Fungi: EuGene as a model platform. . Current Bioinformatics *3*, 87-97.

S24. Degroeve, S., Saeys, Y., De Baets, B., Rouzé, P., and Van de Peer, Y. (2005). SpliceMachine: predicting splice sites from high-dimensional local context representations. Bioinformatics *21*, 1332-1338.

S25. Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. PLoS One *6*, e16526.

S26. Feschotte, C., Keswani, U., Ranganathan, N., Guibotsy, M.L., and Levine, D. (2009). Exploring repetitive DNA landscapes using REPCLASS, a tool that automates the classification of transposable elements in eukaryotic genomes. Genome Biol Evol *1*, 205-220.

S27. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T., et al. (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature *423*, 825-837.

S28. Peters, A.F., van Wijk, S.J., Cho, G.Y., Scornet, D., Hanyuda, T., Kawai, H., Schroeder, D.C., Cock, J.M., and Boo, S.M. (2010). Reinstatement of *E. crouaniorum* Thuret in Le Jolis as a third common species of *Ectocarpus* (Ectocarpales, Phaeophyceae) in western Europe, and its phenology at Roscoff, Brittany. Phycol. Res. *58*, 157-170.

S29. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol *28*, 2731-2739.

S30. Gouy, M., Guindon, S., and Gascuel, O. (2010). SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol *27*, 221-224.

S31. Löytynoja, A., and Goldman, N. (2005). An algorithm for progressive multiple alignment of sequences with insertions. Proc Natl Acad Sci U S A *102*, 10557-10562.

S32. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol *17*, 540-552.

S33. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol *24*, 1586-1591.

S34. Penel, S., Arigon, A.M., Dufayard, J.F., Sertier, A.S., Daubin, V., Duret, L., Gouy, M., and Perrière, G. (2009). Databases of homologous gene families for comparative genomics. BMC Bioinformatics *10 Suppl 6*, S3.

S35. Miele, V., Penel, S., and Duret, L. (2011). Ultra-fast sequence clustering from similarity networks with SiLiX. BMC Bioinformatics *12*, 116.

S36. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol *52*, 696-704.

S37. Bigot, T., Daubin, V., Lassalle, F., and Perrière, G. (2013). TPMS: a set of utilities for querying collections of gene trees. BMC Bioinformatics *14*, 109.

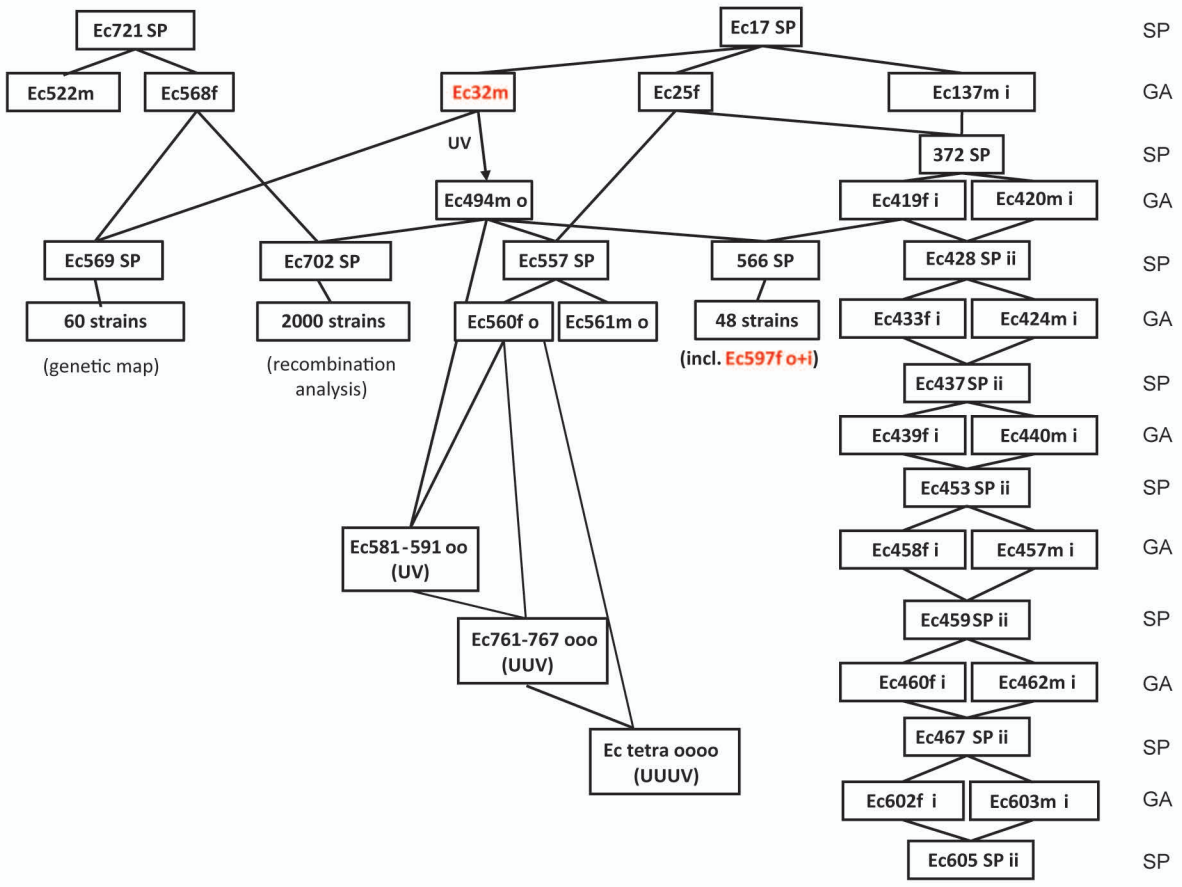
S38. Charif, D., Thioulouse, J., Lobry, J.R., and Perrière, G. (2005). Online synonymous codon usage analyses with the ade4 and seqinR packages. Bioinformatics *21*, 545-547.

S39. Coelho, S.M., Scornet, D., Rousvoal, S., Peters, N., Dartevelle, L., Peters, A.F., and Cock, J.M. (2012). Genetic crosses between *Ectocarpus* strains. Cold Spring Harb Protoc *2012*, 262-265.

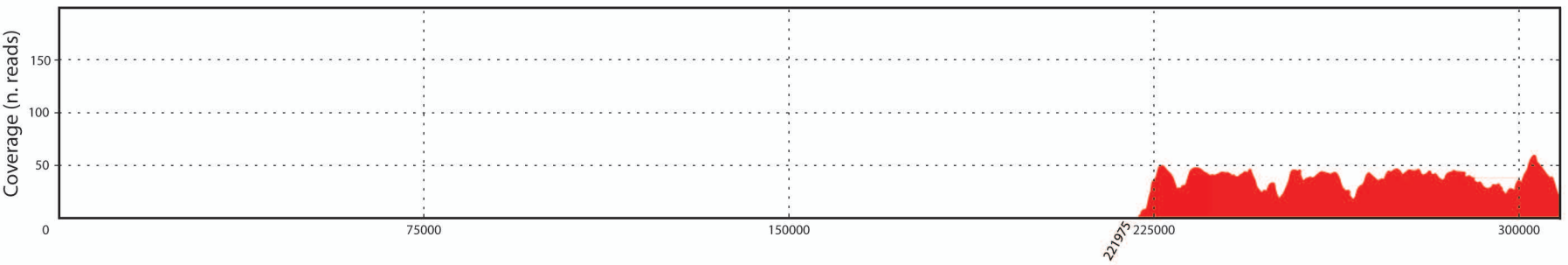
S40. Bothwell, J.H., Marie, D., Peters, A.F., Cock, J.M., and Coelho, S.M. (2010). Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga *Ectocarpus*. New Phytol *188*, 111-121.

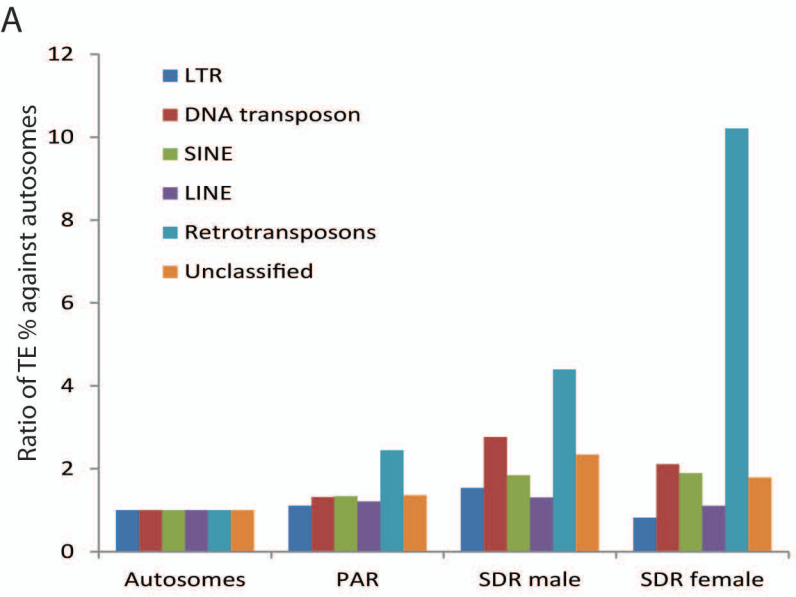
S41. Peters, A.F., Marie, D., Scornet, D., Kloareg, B., and Cock, J.M. (2004). Proposal of *Ectocarpus siliculosus* (Ectocarpales, Phaeophyceae) as a model organism for brown algal genetics and genomics. J Phycol *40*, 1079-1088.

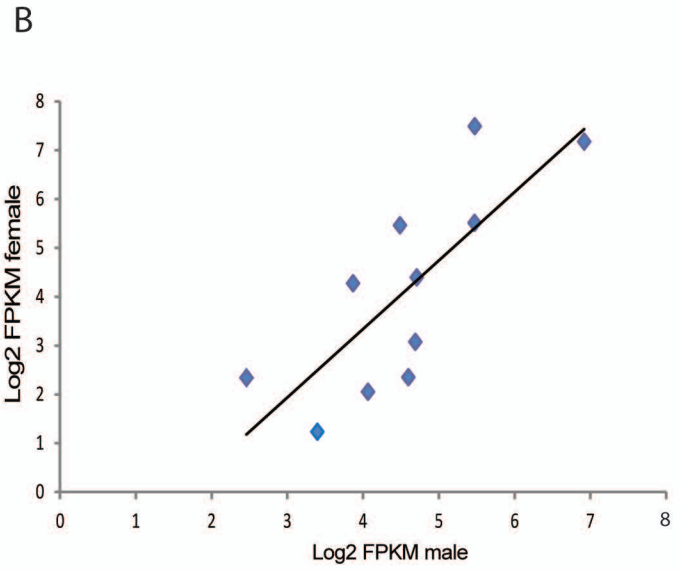
A

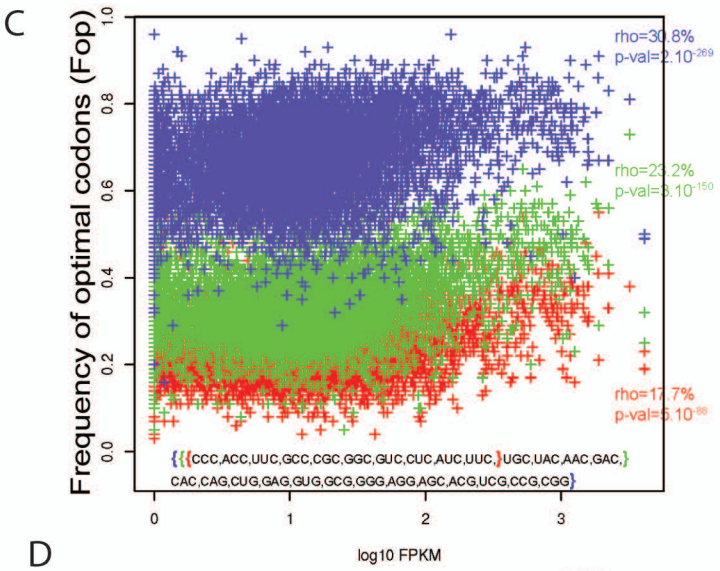

Ec721 SP
Ec17 SP
SP
Ec522m
Ec568f
Ec32m
Ec25f
Ec137m i
GA
UV
372 SP
SP
Ec494m o
Ec419f i
Ec420m i
GA
Ec569 SP
Ec702 SP
Ec557 SP
566 SP
Ec428 SP ii
SP
60 strains
2000 strains
Ec560f o
Ec561m o
48 strains
Ec433f i
Ec424m i
GA
(genetic map)
(recombination analysis)
(incl. Ec597f o+i)
Ec437SP ii
SP
Ec439f i
Ec440m i
GA
Ec453 SP ii
SP
Ec581-591 oo (UV)
Ec458f i
Ec457m i
GA
Ec761-767 ooo (UUV)
Ec459SP ii
SP
Ec460f i
Ec462m i
GA
Ec tetra oooo (UUUV)
Ec467 SP ii
SP
Ec602f i
Ec603m i
GA
Ec605 SP ii
SP

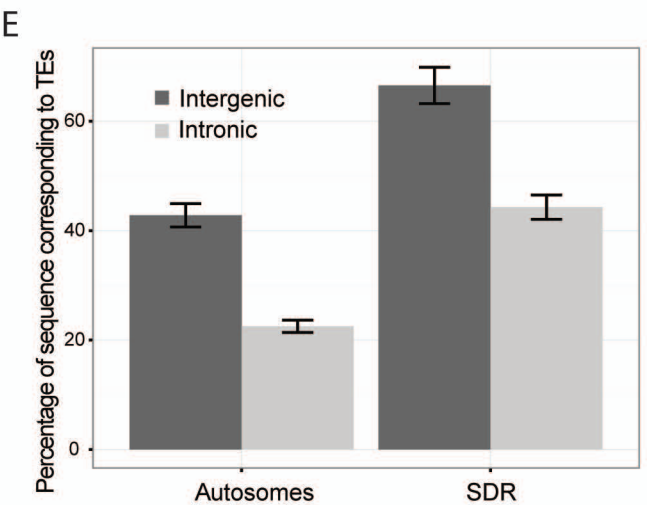
B

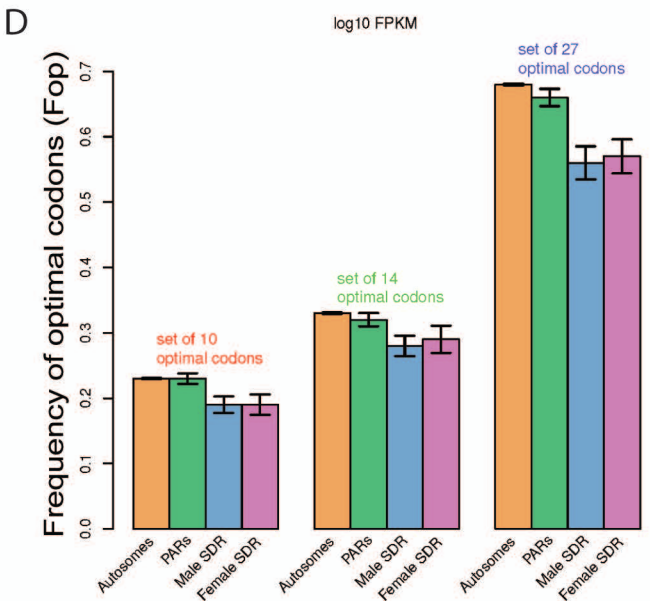

Coverage (n. reads)
150
100
50
0
75000
150000
221975
225000
300000
Position on the sctg439and285

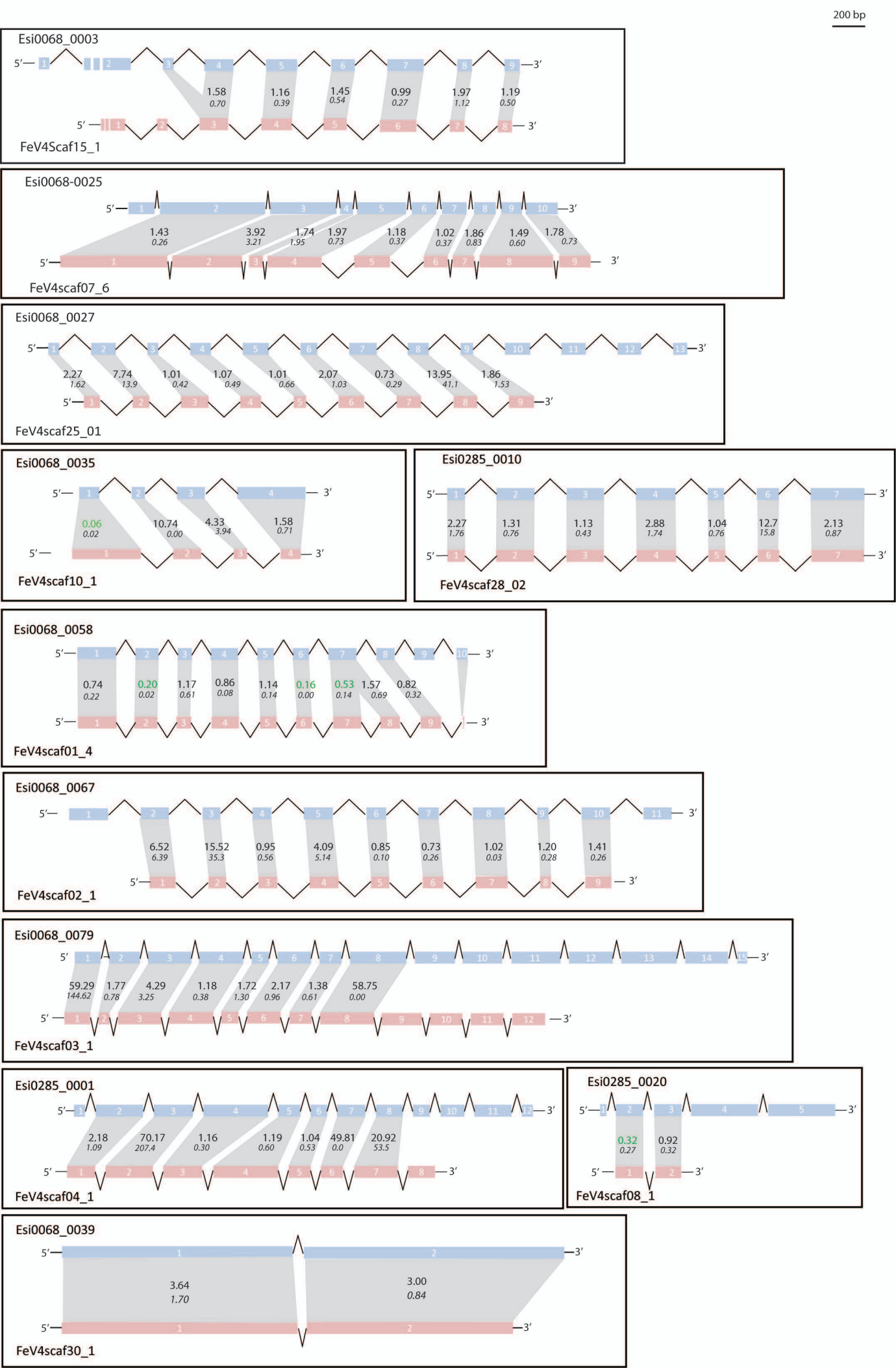

A
Ratio of TE % against autosomes
LTR
DNA transposon
SINE
LINE
Retrotransposons
Unclassified
Autosomes
PAR
SDR male
SDR female


B
Log2 FPKM female
Log2 FPKM male

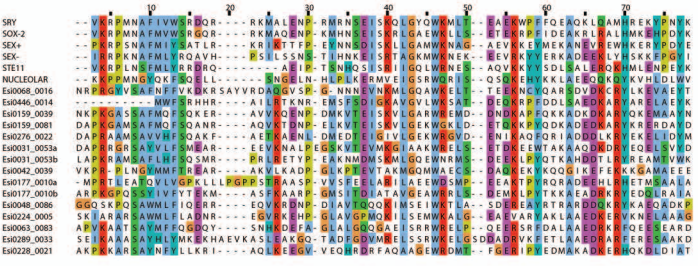

C
Frequency of optimal codons (Fop)
rho=30.8%
p-val=2.10-269
rho=23.2%
p-val=3.10-150
rho=17.7%
p-val=5.10-88
{{CCC,ACC,UUC,GCC,CGC,GGC,GUC,CUC,AUC,UUC,}UGC,UAC,AAC,GAC,}
CAC,CAG,CUG,GAG,GUG,GCG,GGG,AGG,AGC,ACG,UCG,CCG,CGG}
log10 FPKM


E
Percentage of sequence corresponding to TEs
Intergenic
Intronic
Autosomes
SDR

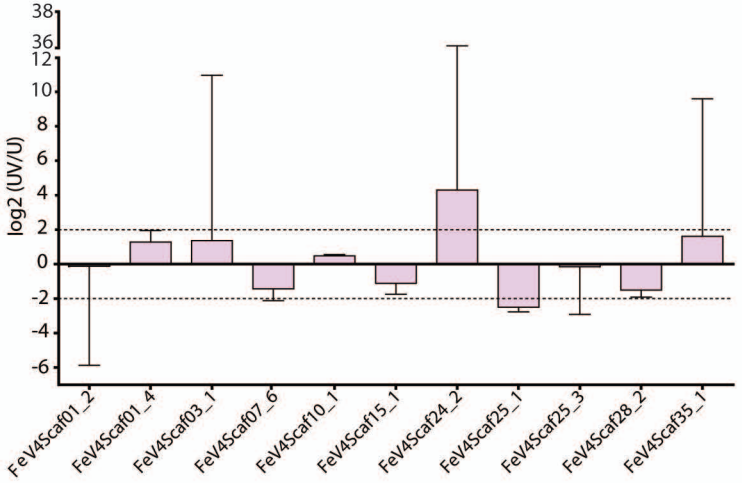

D
Frequency of optimal codons (Fop)
set of 10
optimal codons
set of 14
optimal codons
set of 27
optimal codons
Autosomes
PARs
Male SDR
Female SDR


200 bp
Esi0068_0003
FeV4Scaf15_1
Esi0068-0025
FeV4scaf07_6
Esi0068_0027
FeV4scaf25_01
Esi0068_0035
FeV4scaf10_1
Esi0285_0010
FeV4scaf28_02
Esi0068_0058
FeV4scaf01_4
Esi0068_0067
FeV4scaf02_1
Esi0068_0079
FeV4scaf03_1
Esi0285_0001
FeV4scaf04_1
Esi0285_0020
FeV4scaf08_1
Esi0068_0039
FeV4scaf30_1

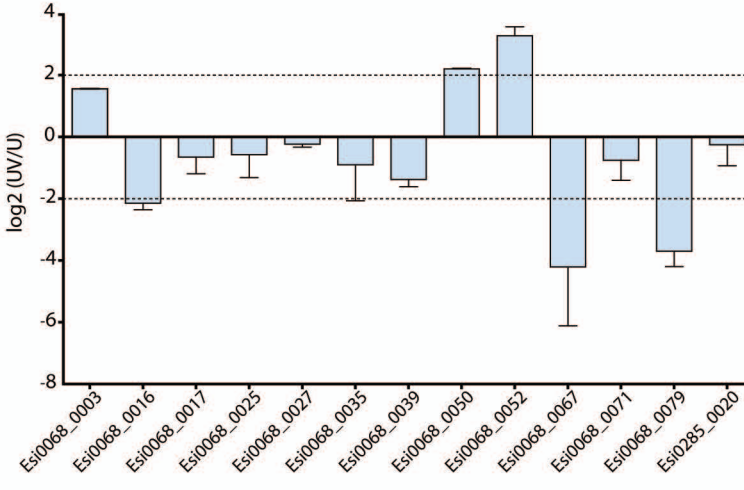
A

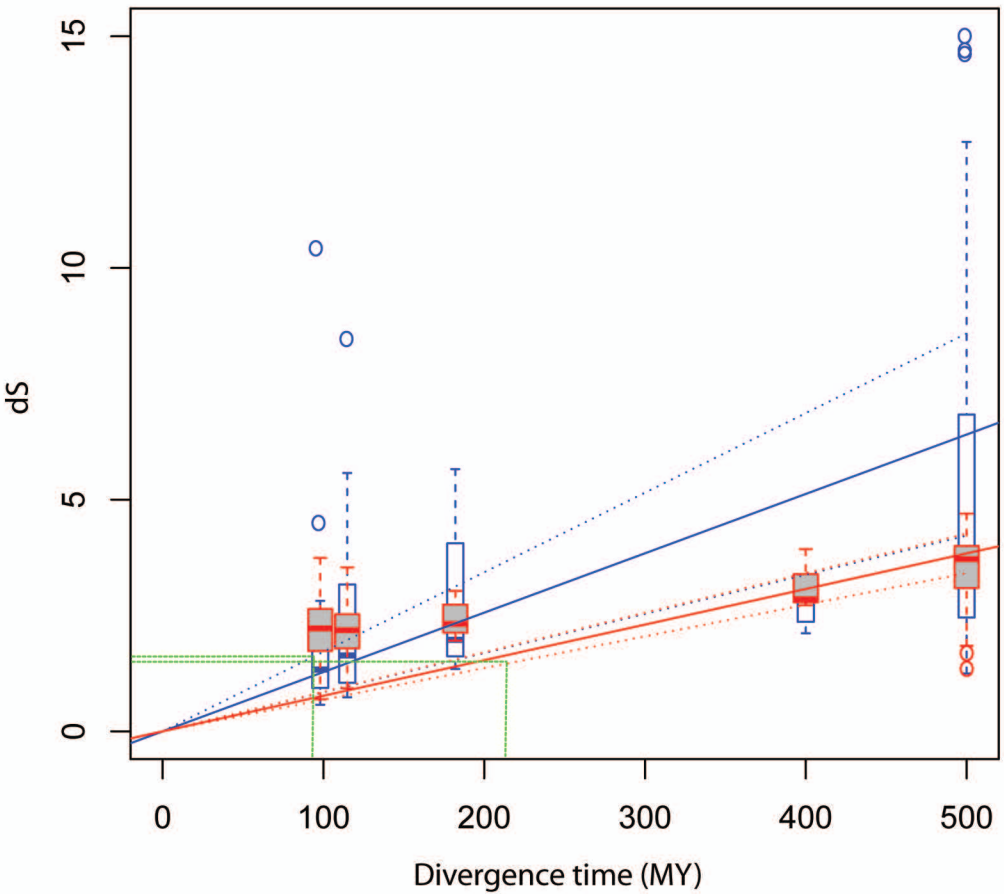

SRY
SOX-2
SEX+
SEX-
STE11
NUCLEOLAR
Esi0068_0016
Esi0446_0014
Esi0159_0039
Esi0159_0081
Esi0276_0022
Esi0031_0053a
Esi0031_0053b
Esi0042_0039
Esi0177_0010a
Esi0177_0010b
Esi0048_0086
Esi0224_0005
Esi0063_0083
Esi0289_0033
Esi0228_0021

B


log2 (UV/U)
FeV4Scaf01_2
FeV4Scaf01_4
FeV4Scaf03_1
FeV4Scaf07_6
FeV4Scaf10_1
FeV4Scaf15_1
FeV4Scaf24_2
FeV4Scaf25_1
FeV4Scaf25_3
FeV4Scaf28_2
FeV4Scaf35_1

C


log2 (UV/U)
Esi0068_0003
Esi0068_0016
Esi0068_0017
Esi0068_0025
Esi0068_0027
Esi0068_0035
Esi0068_0039
Esi0068_0050
Esi0068_0052
Esi0068_0067
Esi0068_0071
Esi0068_0079
Esi0285_0020


dS
0
5
10
15
0
100
200
300
400
500
Divergence time (MY)