Supplementary information for
Genes and the species concept -
How much of the genomes can be exchanged?

Xinfeng Wang[a, *], Zixiao Guo[a, *], Shaohua Xu[a], Ming Yang[a], Qipian Chen[a], Sen Li[a], Cairong Zhong[b], Norman C. Duke[c], Ziwen He[a], Chung-I Wu[a,d,e] & Suhua Shi[a]

[a] State Key Laboratory of Biocontrol, Guangdong Key Lab of Plant Resources, Key Laboratory of Biodiversity Dynamics and Conservation of Guangdong Higher Education Institutes, School of Life Sciences, Sun Yat-Sen University, Guangdong, China
[b] Hainan Dongzhai Harbor National Nature Reserve Administration, Haikou, China
[c] Centre for Tropical Water and Aquatic Ecosystem Research, James Cook University, Townsville, Australia
[d] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
[e] Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA
[*] These authors contributed equally to this work
Correspondence should be addressed to S.S. (lssssh@mail.sysu.edu.cn), C.-I.W. (ciwu@uchicago.edu) or Z.H. (heziwen@mail.sysu.edu.cn)

This file includes:

Supplementary Notes

In all figures and tables, the *R. mucronata* populations are labeled as m1, m2, m3, m4, m5, m6, and m7, while *R. stylosa* populations are labeled as s1, s2, s3, and s4, where "m" stands for *R. mucronata* and "s" stands for *R. stylosa* (Table S1). To refer to all the *R. stylosa* or *R. mucronata* populations, we use "S_{all}" or "M_{all}". We use "S_{allo}" to refer to the allopatric *R. stylosa* populations s2-s4 and "M_{allo}" to the allopatric *R. mucronata* populations m2-m7.

Additional introgression tests between sympatric species

i) When comparing the two species, we identified 11,756 fixed SNPs and 759,472 shared SNPs (Table S5). Removing m1 and s1 (or DR samples), changes the SNP counts to 194,172 and 268,132 respectively. Removing allopatric populations does not affect the locus counts much, and the numbers remain around 12,000 and 720,000 (Table S5). This suggests that incomplete lineage sorting as not the cause of the observed admixture in the DR samples.

ii) We also observed an increased genome-wide linkage disequilibrium (LD) in m1 and s1 (Supplementary Fig. S5). Hence, if the background admixture in the allopatric populations reflects ancient polymorphism (or incomplete lineage sorting), the excess admixture we observed in m1 and s1 can be reasonably attributed to introgression private to m1 and s1.

iii) We further used the ABBA-BABA (or *D*) statistic to test for the excess of shared derived alleles due to introgression[76,77]. A positive "*D*" or "f_d" value is an indicator of introgression (gene flow). The tests with m1 and s1 as the subject branches all showed significant positive "*D*" and "f_d" values ($P<0.01$, Supplementary Table S6). In contrast, no mean *D* values significantly deviated from 0 ($P>0.5$) in the tests without m1 or s1 (Supplementary Table S6). Using 500kb sliding windows, we constructed distributions of *D* values across the genome. Between 98.17% and 99.48% of the windows had positive *D* values in the tests with m1 and s1 (Supplementary Fig. S6). In other words, introgression was only found between m1 and s1 locally in Daintree River, where *R. stylosa* and *R. mucronata* are found together.

Detecting highly differentiated amino acids between R. mucronata and R. stylosa

To find highly differentiated amino acids in the 30 candidate genes in j-blocks, we obtained protein sequences of all 52 individuals for each gene (Table 2 and Supplementary Table S10). We used the following criteria to call highly differentiated amino acids: 1) the differentiated nucleotide in the codon is a non-synonymous site and with $F_{ST} > 0.8$ between M_{allo} and S_{allo} samples; 2) there are no identical homozygotes between *R. mucronata* and *R. stylosa*. We found 30 such amino acids between *R. mucronata* and *R. stylosa* in the 30 genes (Supplementary Table S10). Fourteen of these sites are located in the seven genes involved in flower development (Table 2 and Supplementary Fig. S12).

The possible evolutionary trajectory and speciation of R. mucronata and R. stylosa

These observations re-enforce the notion that species characters vary over geographic range. Accepting that, we then ask what additional circumstances would be needed for speciation to occur and how might these conditions appear for the two *Rhizophora* species in this case study?

The characters shown for the two closely related species by these genetic and morphological studies have revealed distinct patterns and traits that appear to follow an ordered series of features tending towards genetic isolation and speciation. The order of events may have followed a scenario similar to the following[37]:

1) It started with the dispersal of Asian *R. mucronata* into a vacant habitat as conditions became suitable in the west and south, but with concurrent selection favoring drift towards arid and marine traits to differentiate East African *R. mucronata*;
2) Propagules were transported by their exceptional ability for long distance dispersal and a specialized capability for establishment and growth in marine coastal conditions in a broad range of wet and dry climatic conditions;

3) These circumstances would have transported populations south and east towards Australia (in one or more founder events that selected a subset of genotypes which at some point became *R. stylosa*);
4) The longer style of *R. stylosa* further implies some possibly greater reliance on a particular pollinator, but this has not been established;
5) Ancestral Australian *R. stylosa* migrates further east and north along northern Australian shorelines, and then into SE Asia as well as the western Pacific where at some point it re-unites with ancestral Asian *R. mucronata* (still ecologically conditioned for upstream brackish estuarine locations);
6) Before this event, Asian *R. mucronata* populations would have been expanding east and south as conditions became more suitable across the region, but the proximity and size of populations would likely have prevented significant isolation events and no further speciation would have occurred.
7) The geomorphic circumstances of continental drift and their timing were likely critical in the progress of the biological events of dispersal and speciation[79].

Supplementary Tables

Table S1 Sampling, re-sequencing information, and genetic diversity statistics of *R. mucronata* and *R. stylosa* populations.

Sampling location	Longitude, latitude	Pop ID	Sample size	Effective sites [i]	Coverage rate [ii]	SNPs ($x10^5$)	θ_w/Kb	θ_π/Kb
R. mucronata								
Daintree River, Australia	145 °26'16.28" E,16 °17'12.44" S	m1	5	1.90E+08	0.82	5.65	1.06	1.05
Saint John's Island, Singapore	103 °50'30.19" E, 1 °13'6.60" N	m2	5	1.89E+08	0.81	2.41	0.53	0.45
Chai-Ya, Thailand	99 °15'32.35" E, 9 °21'23.07" N	m3	4	1.86E+08	0.80	1.72	0.41	0.35
Ranong, Thailand	98 °37'26.01" E, 9 °57'36.26" N	m4	4	1.88E+08	0.81	3.80	0.89	0.79
Tanjung Piai, Malaysia	103 °21'1.86" E, 1 °24'8.11" N	m5	5	1.90E+08	0.82	2.84	0.60	0.53
Mauritius	57 °41'18.33" E, 20 °20'26.37" S	m6	3	1.84E+08	0.79	0.81	0.22	0.19
Kenya	39 °36'1.82" E, 4 °24'24.15" S	m7	5	1.83E+08	0.79	1.55	0.35	0.30
R. stylosa								
Daintree River, Australia	145 °26'16.28" E, 16 °17'12.44" S	s1	5	1.88E+08	0.81	6.47	1.52	1.22
Saint John's Island, Singapore	103 °50'30.19" E, 1 °13'6.60" N	s2	5	1.87E+08	0.81	1.31	0.28	0.25
Daiwin, Australia	130 °54'22.64" E, 12 °25'5.75" S	s3	6	1.87E+08	0.81	4.19	0.86	0.74
Hainan, China	110 °35'5.79" E, 19 °56'39.67" N	s4	5	1.85E+08	0.79	1.07	0.24	0.21

[i] Effective sites: the sites that have at least one individual mapped (depth>=2) for each population.

[ii] Coverage rate: the ratio of effective sites and reference genome size. The reference (*R. apiculata*) genome size equals to 232,430,847 bp.

Table S2 Re-sequencing characteristics and heterozygosity of each *R. mucronata* genome

Population ID	Individual ID	Read length (bp)	Raw read pairs (E+07)	Retained read pairs (E+07)	Mapped reads (E+07)	Reads mapping rate (%)	Effective sites [i] (E+08)	Coverage rate [ii]	Mean Depth [iii]	Ht[iv] (per Kb)
m1 (Daintree River, Australia)	m1-1	125	2.03	2.03	3.22	79.52	1.79	0.77	17X	1.430
	m1-15	125	1.37	1.37	2.26	82.29	1.54	0.66	12X	0.269
	m1-16	125	1.36	1.36	2.18	79.95	1.75	0.75	12X	1.509
	m1-2	100	2.18	1.99	3.46	86.93	1.79	0.77	15X	0.369
	m1-3	100	2.51	2.28	3.96	86.77	1.80	0.78	17X	0.312
m2 (Saint John's Island, Singapore)	m2-1	150	1.49	1.43	2.43	85.14	1.76	0.76	16X	0.606
	m2-2	150	2.09	1.98	3.36	84.79	1.79	0.77	22X	0.378
	m2-3	150	1.62	1.58	2.84	89.94	1.80	0.77	18X	0.458
	m2-4	150	0.14	1.40	2.52	90.15	1.82	0.78	16X	0.483
	m2-5	150	1.78	1.73	3.10	89.51	1.80	0.77	20X	0.533
m3 (Chai-Ya, Thailand)	m3-12	125	1.48	1.48	2.58	87.00	1.79	0.77	14X	0.351
	m3-14	125	1.62	1.62	2.85	88.16	1.81	0.78	15X	0.374
	m3-5	125	1.51	1.51	2.68	88.52	1.80	0.78	14X	0.362
	m3-7	125	1.42	1.42	2.49	87.77	1.79	0.77	13X	0.381
m4 (Ranong, Thailand)	m4-1	125	1.70	1.70	2.97	87.37	1.82	0.78	16X	0.351
	m4-2	125	1.46	1.46	2.57	88.02	1.81	0.78	14X	0.475
	m4-3	125	1.65	1.65	2.93	88.46	1.82	0.78	16X	0.461
	m4-4	125	1.49	1.49	2.61	87.77	1.82	0.78	14X	0.514
m5 (Tanjung Piai, Malaysia)	m5-12	125	1.40	1.34	2.34	87.24	1.79	0.77	13X	0.656
	m5-14	125	1.90	1.90	2.97	77.96	1.79	0.77	16X	0.340
	m5-21	125	2.01	2.01	3.22	79.82	1.80	0.77	17X	0.449
	m5-32	125	1.34	1.28	2.24	87.44	1.78	0.77	12X	0.567
	m5-9	125	1.27	1.19	2.07	87.30	1.75	0.75	11X	0.369
m6 (Mauritius)	m6-1	125	1.35	1.35	2.34	86.72	1.77	0.76	13X	0.296
	m6-2	125	1.34	1.34	2.36	88.19	1.77	0.76	13X	0.266
	m6-3	125	1.32	1.32	2.28	86.46	1.76	0.76	12X	0.280
m7 (Kenya)	m7-10	125	1.55	1.55	2.51	80.99	1.75	0.75	13X	0.389
	m7-11	125	1.48	1.48	2.36	79.54	1.74	0.75	13X	0.386
	m7-2	125	1.77	1.77	2.80	78.81	1.76	0.76	15X	0.317
	m7-7	125	2.11	2.11	3.40	80.31	1.78	0.77	18X	0.414
	m7-9	125	1.92	1.92	3.08	80.06	1.76	0.76	17X	0.411

[i] Effective sites: all sites that have at least two reads mapped (depth>=2) in each site.

[ii] Coverage rate: the ratio of effective sites and reference genome size. The reference genome size equals to 232430847 bps.

[iii] Mean depth: the average number of reads mapped at those sites.

[iv] Mean genome-wide heterozygosity (Ht) : the proportion occupied by heterozygotes of the genome.

Table S3 Re-sequencing characteristics and heterozygosity of each *R. stylosa* genome

Population ID	Individual ID	Read length (bp)	Raw read pairs (E+07)	Retained read pairs (E+07)	Mapped reads (E+07)	Reads mapping rate (%)	Effective sites [i] (E+08)	Coverage rate [ii]	Mean Depth [iii]	Ht [iv] (per Kb)
s1 (Daintree River, Australia)	**s1-10**	100	1.83	1.68	2.90	86.34	1.81	0.78	12X	2.264
	s1-5	100	1.72	1.58	2.74	86.78	1.77	0.76	12X	1.109
	s1-6	100	1.95	1.79	3.08	86.20	1.79	0.77	13X	1.395
	s1-7	100	2.01	1.85	3.20	86.73	1.80	0.78	14X	2.288
	s1-9	100	1.81	1.66	2.89	86.82	1.78	0.77	12X	1.386
m2 (Saint John's Island, Singapore)	**s2-1**	150	1.58	1.54	2.73	89.01	1.78	0.77	18X	0.327
	s2-3	150	1.72	1.68	3.01	89.77	1.81	0.78	19X	0.297
	s2-4	150	1.53	1.48	2.64	89.18	1.77	0.76	17X	0.320
	s2-5	150	1.41	1.38	2.46	88.97	1.79	0.77	16X	0.298
	s2-6	150	1.43	1.39	2.51	89.95	1.79	0.77	16X	0.308
s3 (Darwin, Australia)	**s3-1**	100	2.24	2.07	3.56	86.11	1.79	0.77	15X	0.826
	s3-2	100	2.19	2.01	3.40	84.69	1.79	0.77	15X	0.795
	s3-3	100	1.94	1.80	3.08	85.70	1.76	0.76	13X	0.532
	s3-4	100	2.16	1.97	3.30	83.51	1.78	0.77	14X	0.782
	s3-5	100	2.26	2.07	3.54	85.57	1.79	0.77	15X	0.822
	s3-6	100	2.69	2.46	4.21	85.60	1.81	0.78	18X	0.869
s4 (Hainan, China)	**s4-10**	100	1.99	1.82	3.14	86.46	1.77	0.76	14X	0.288
	s4-13	100	2.00	1.84	3.14	85.21	1.79	0.77	14X	0.271
	s4-3	100	1.85	1.69	2.92	86.17	1.77	0.76	13X	0.270
	s4-4	100	1.79	1.65	2.85	86.70	1.77	0.76	12X	0.270
	s4-7	100	1.91	1.75	3.04	86.65	1.77	0.76	13X	0.265

[i] Effective sites: all sites that have at least two reads mapped (depth>=2) in each site.

[ii] Coverage rate: the ratio of effective sites and reference genome size. The reference genome size equals to 232430847 bps.

[iii] Mean depth: the average number of reads mapped at those sites.

[iv] Mean genome-wide heterozygosity (Ht) : the proportion occupied by heterozygotes of the genome.

Table S4 Additional diagnostic morphological features to identify *R. mucronata* and *R. stylosa*

Feature	***R. mucronata***	***R. stylosa***
bracts and bracteoles	minute bracts and bracteoles	distinct bracts and bracteoles
inflorescences	1-2 flowered inflorescences	4-16 flowered inflorescences
flower buds	irregular obovoid closed flower buds	regular ovoid-elliptic closed flower buds
propagules	long propagules reaching ~80 cm	~60 cm

Table S5 Information on genomic polymorphisms in *R. mucronata* and *R. stylosa*

Removed populations	**Fixed difference**	**Shared polymorphisms**	**Private polymorphisms in *R. stylosa***	**Private polymorphisms in *R. mucronata***	**Total SNPs**
none	11756	759472	227815	460472	1459515
m1, s1	194172	268132	343000	501487	1306791
m2, s2	12748	717935	216114	458777	1405574
m2, s3	12731	695504	169470	487094	1364799
m2, s4	12834	731540	223343	453017	1420734
m3, s2	12566	721858	212522	476685	1423631
m3, s3	11584	705580	160213	500929	1378306
m3, s4	11734	741871	213954	466712	1434271
m4, s2	12906	716301	217948	436973	1384128
m4, s3	11944	700769	164884	460400	1337997
m4, s4	12161	736081	219576	427083	1394901
m5, s2	12503	718950	215390	460276	1407119
m5, s3	11562	702866	162876	483706	1361010
m5, s4	11758	738800	216928	449683	1417169
m6, s2	12764	720055	214363	472138	1419320
m6, s3	11877	704025	161756	495947	1373605
m6, s4	12074	739901	215921	462133	1430029
m7, s2	12583	716611	217785	460798	1407777
m7, s3	11672	701937	163854	482878	1360341
m7, s4	11850	736356	219454	450360	1418020

Removed populations: we removed two populations from all samples each time and then calculate the polymorphisms in the retained *R. mucronata* and *R. stylosa*. "none" means we kept all samples.

Table S6 Patterson's *D* statistic and improved f_d statistic, showing evidence of gene flow between *R. mucronata* (m1) and *R. stylosa* (s1) in sympatry in Daintree River, Australia

Model code [a]	Pop1 [b]	Pop2 [b]	Pop3 [b]	Outgroup [b]	*D* ± std err [c]	Z-score	P-value	f_d ± std err [d]
1	m7	m6	s1	ra	0.000980 ±0.0135	0.00372	0.997	0.000313 ±0.000257
2	m7	m6	s2	ra	-0.0160 ±0.0197	-0.0414	0.967	0.00535 ±0.000261
3	m7	m6	s3	ra	0.0117 ±0.0190	0.0315	0.975	0.000300 ±0.000207
4	m7	m6	s4	ra	0.00201 ±0.0199	0.00516	0.996	0.00388 ±0.000265
5	m3	m2	s2	ra	-0.131 ±0.0122	-0.548	0.584	-0.0148 ±0.000213
6	m4	m2	s2	ra	0.0750 ±0.0133	0.288	0.773	0.0148 ±0.000439
7	m5	m2	s2	ra	-0.112 ±0.0100	-0.568	0.570	-0.00486 ±0.000165
8	m6	m2	s2	ra	-0.203 ±0.0187	-0.555	0.579	-0.0414 ±0.000557
9	m7	m2	s2	ra	-0.197 ±0.0191	-0.528	0.597	-0.0370 ±0.000775
10	s3	s2	m2	ra	0.116 ±0.0170	0.349	0.727	0.0630 ±0.00239
11	s4	s2	m2	ra	-0.114 ±0.184	-0.318	0.750	-0.0239 ±0.00135
12	m2	m1	s1	ra	0.637 ±0.00807	4.04***	**5.35E-05**	0.482 ±0.000300
13	m3	m1	s1	ra	0.628 ±0.00867	3.70***	**2.16E-04**	0.470 ±0.000270
14	m4	m1	s1	ra	0.649 ±0.00859	3.86***	**1.13E-04**	0.483 ±0.000297
15	m5	m1	s1	ra	0.630 ±0.00826	3.90***	**9.62E-05**	0.582 ±0.000183
16	m6	m1	s1	ra	0.556 ±0.0101	2.81***	**4.95E-03**	0.454 ±0.000325
17	m7	m1	s1	ra	0.555 ±0.0101	2.81***	**4.95E-03**	0.455 ±0.000349
18	s0	s1	m1	ra	0.580 ±0.0104	2.86***	**4.24E-03**	0.543 ±0.000345
19	s3	s1	m1	ra	0.592 ±0.00931	3.25***	**1.15E-03**	0.552 ±0.000323
20	s4	s1	m1	ra	0.572 ±0.0102	2.87***	**4.10E-03**	0.533 ±0.000314

[a]: code of *D* statistic models. Models 1-11 exclude sympatric populations m1 and s1; models 12-20 contain sympatric populations m1 and s1.

[b]: Pop1, Pop2, Pop3 and Outgroup respectively refer to the three ingroups and outgroup (ra: *R. apiculata*) followed the genealogical relationship (((Pop1, Pop2), Pop3), Outgroup).

[c]: result of *D* statistic, given as a ratio *D* ± standard error.

***: the genome-wide average *D*-statistic value *D* is significantly derived from 0 with $P < 0.01$, indicating the existence of gene flow between m1 and s1 populations.

[d]: result of f_d statistic, given as a admixed proportion f_d ± standard error.

Table S7 The introgressed site (i-site) distribution across introgressed blocks (or i-blocks) in m1 and s1 genomes

The i-sites range of i-blocks	>=2 occurrences of i-allele		>=4 occurrences of i-allele		>=6 occurrences of i-allele		>=8 occurrences of i-allele		=10 occurrences of i-allele	
	m1	**s1**	**m1**	**s1**	**m1**		**m1**	**s1**	**m1**	**s1**
1 (singleton block)	13499	7500	12708	6659	12580	5782	12586	4319	12182	2746
2	3827	2282	3805	2140	3830	1736	3791	1243	3548	694
3	1624	1143	1648	1114	1662	837	1656	563	1486	297
4	749	673	759	669	780	558	761	325	694	202
5	401	463	420	458	423	361	428	241	401	133
5-10	657	840	678	857	691	687	686	434	625	219
10-15	186	208	193	211	194	195	186	130	171	67
15-20	87	79	87	88	87	88	86	55	60	26
20-30	72	58	73	61	74	60	75	47	44	29
30-40	27	21	28	18	27	18	25	16	9	12
40-50	10	10	11	10	11	13	6	8	5	5
50-60	2	3	2	3	2	4	2	5	1	4
60-70	0	2	0	2	0	1	1	0	0	0
70-80	5	0	5	0	5	0	5	0	1	0
80-90	3	3	3	3	3	3	2	2	1	1
90-100	2	0	2	0	2	0	2	0	0	0
>100	2	1	3	1	3	1	2	1	0	0
Total blocks	21153	13286	20425	12294	20374	10344	20300	7389	19228	4435
Blocks (>=2 i-sites)	7654	5786	7717	5635	7794	4562	7714	3070	7046	1689

Table S8 The length distribution of introgressed blocks (i-blocks) in m1 and s1 genomes

Length range of i-blocks	>=2 occurrences of i-allele		>=4 occurrences of i-allele		>=6 occurrences of i-allele		>=8 occurrences of i-allele		=10 occurrences of i-allele	
	m1	**s1**	**m1**	**s1**	**m1**		**m1**	**s1**	**m1**	**s1**
1-10bp	48	26	47	22	47	13	47	11	45	6
10-100bp	2150	1078	2095	945	2084	822	2083	604	2019	322
100bp-1Kb	12230	6966	11660	6216	11607	5165	11637	3558	11205	2113
1Kb-5Kb	5329	4018	5218	3876	5213	3217	5163	2338	4832	1427
5Kb-10Kb	734	664	738	690	749	595	731	458	643	290
10Kb-20Kb	373	351	384	350	388	344	367	279	298	185
20Kb-30Kb	133	83	130	84	133	86	126	72	94	46
30Kb-50Kb	89	69	87	75	86	66	84	48	47	33
50Kb-100Kb	43	26	42	32	43	31	40	19	30	11
>100Kb	**24**	**5**	**24**	**4**	**24**	**5**	**22**	**2**	**15**	**2**
Total blocks	21153	13286	20425	12294	20374	10344	20300	7389	19228	4435

Table S9 Detailed information on introgressed blocks (i-blocks) in m1 and s1 genomes

Description		>=2 occurrences of i-allele		>=4 occurrences of i-allele		>=6 occurrences of i-allele		>=8 occurrences of i-allele		=10 occurrences of i-allele	
		m1 pop	**s1 pop**	**m1 pop**	**s1 pop**	**m1 pop**		**m1 pop**	**s1 pop**	**m1 pop**	**s1 pop**
The i-blocks contain >=1 intro sites	No. of i-blocks	21153	13286	20425	12294	20374	10344	20300	7389	19228	4435
	No. of scaffolds with i-blocks	99	98	99	99	99	97	99	97	99	95
	Total length of i-blocks (bp)	40,335,353	28,733,479	39,888,762	28,759,674	40,135,421	25,519,942	38,687,939	18,749,939	32,136,536	11,918,934
	% of the genome	24.355	17.350	24.086	17.366	24.235	15.409	23.361	11.322	19.405	7.197
The i-blocks contain >=2 intro sites	No. of i-blocks	7654	5786	7717	5635	7794	4562	7714	3070	7046	1689
	No. of scaffolds with i-blocks	96	97	96	96	96	96	96	93	96	88
	Total length of i-blocks (bp)	26,812,491	19,881,117	27,387,055	20,624,605	27,716,879	18,085,016	26,181,467	12,917,006	19,686,486	7,768,779
	% of the genome	16.190	12.005	16.537	12.454	16.736	10.920	15.809	7.800	11.887	4.691
The i-blocks contain >=3 intro sites	No. of i-blocks	3827	3504	3912	3495	3964	2826	3923	1827	3498	995
	No. of scaffolds with i-blocks	96	97	96	96	96	93	96	91	95	84
	Total length of i-blocks (bp)	19,593,946	15,639,138	20,042,145	16,341,288	20,313,299	14,507,460	19,097,627	10,112,925	13,410,594	6,100,153
	% of the genome	11.831	9.443	12.102	9.867	12.266	8.760	11.532	6.106	8.098	3.683
The i-blocks contain >=4 intro sites	No. of i-blocks	2203	2361	2264	1381	2302	1989	2267	1264	2012	698
	No. of scaffolds with i-blocks	96	90	96	91	96	88	96	88	95	81
	Total length of i-blocks (bp)	14,800,743	12,650,069	15,208,066	13,210,483	15,448,953	11,967,003	14,214,177	8,238,250	9,636,550	4,944,119
	% of the genome	8.937	7.638	9.183	7.977	9.328	7.226	8.583	4.974	5.818	2.985
The i-blocks contain >=5 intro sites	No. of i-blocks	1454	1688	1505	1712	1522	1431	1506	939	1318	496
	No. of scaffolds with i-blocks	96	90	96	91	96	88	96	85	95	79
	Total length of i-blocks (bp)	12,428,169	10,694,031	12,792,450	11,033,783	12,967,704	10,187,524	11,663,533	6,941,326	7,474,717	4,021,295
	% of the genome	7.504	6.457	7.724	6.662	7.830	6.151	7.043	4.191	4.513	2.428

Table S10 All functional genes within non-introgressable blocks (j-blocks) between *R. mucronata* and *R. stylosa*

In *Rhizophora*				In *Arabidopsis thaliana*	
Gene	**j-sites**	**L(aa)**	**Mutant**	**Gene**	**Function**
RA_05397	1	85	0	*AT4G31940*	Encodes a cytochrome P450 enzyme, CYP82C. It is involved in the early Fe deficiency response. Other names: CYP82C4.
RA_07374	2	145	0	*AT4G26370*	RNA binding domain of NusB (N protein-Utilization Substance B). The NusB protein plays a key role in the regulation of ribosomal RNA biosynthesis in eubacteria by modulating the efficiency of transcriptional antitermination.
RA_07376	2	172	0	*AT1G55490*	Encodes the beta subunit of the chloroplast chaperonin 60, a homologue of bacterial GroEL. Mutants in this gene develops lesions on its leaves, expresses systemic acquired resistance (SAR) and develops accelerated cell death to heat shock stress. Other names: CPN60B, CPN60BETA1, CPNB1, LEN1. This gene can participate in embryo and seed development[80,81].
RA_08385	2	286	1	*AT4G31050*	Biotin/lipoate A/B protein ligase family. Involved in: protein modification process, lipoate biosynthetic process; Located in: cytoplasm. Other names: LIP2, OCTANOYL- TRANSFERASE.
RA_08689	**2**	**259**	**2**	***AT4G37280***	**MRG family protein. Regulating flowering through elevating the expression of flowering genes *FLC* and *FT* (FLOWERING LOCUS C and T). The mutant shows a late-flowering phenotype. Involved in pollen germination, tube growth and cotyledon development.**
RA_08699	**1**	**452**	**1**	***AT3G22990***	**Armadillo-repeat containing protein. Other name: LEAF AND FLOWER RELATED, *LFR*. Required for all stages of pollen development. The expression is particularly strong in the tapetal cells and pollen grains. The null allele is male-sterile.**
RA_08760	2	1105	2	*AT1G31690*	Copper amine oxidase family protein. Involved in: amine metabolic process. Expressed during leaf development.
RA_08805	1	289	0	*AT3G06200*	P-loop containing nucleoside triphosphate hydrolases superfamily protein. Involved in chloroplast,cytosol phosphorylation, regulation of developmental growth. Located in chloroplast and cytosol.
RA_08848	2	560	2	*AT2G39090*	Tetratricopeptide repeat (TPR)-containing protein. Other names: APC7, ATAPC7. Involved in cell cycle, cell division,protein ubiquitination. Located in nucleus.
RA_10417	**1**	**461**	**3**	***AT4G32440***	**Plant Tudor-like RNA-binding protein. Involved in biological progress. Located in nucleus. Participating in pollen germination and tube growth.**
RA_11619	**2**	**300**	**2**	***AT5G17200***	**Pectin lyase-like superfamily protein. Involved in carbohydrate metabolic process. Located in endomembrane system. Participating in early stage of female gametophyte development.**

RA_13641	**2**	**551**	**1**	***AT3G56600***	**Phosphatidylinositol 4-kinase gamma-like protein. Involved in phosphorylation. Expressed during: L mature pollen stage, M germinated pollen stage, 4 anthesis, petal differentiation and expansion stage. Expressed in flower tissues. Participating in pollen germination and tube growth.**
RA_15569	2	754	1	*AT1G08520*	Encodes the CHLD subunit of the Mg-chelatase enzyme. Involved in chlorophyll biosynthesis. Located in chloroplast and extracellular regions. Participating in embryo and seed development[82]. Lines carrying recessive mutations of this locus are white and seedling lethal. Other names: ALB-1V, ALB1, ALBINA 1, CHLD, PDE166, PIGMENT DEFECTIVE EMBRYO 166, V157.
RA_15894	2	560	0	*AT3G23310*	AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein. Involved in protein phosphorylation. Located in cytosol, nucleus and plasma membrane.
RA_16076	2	1162	3	*AT2G30070*	Encodes a high affinity potassium transporter. Involved in cell tip growth, potassium ion transmembrane transport and potassium ion transport. Located in membrane. Other name: ATKT1, ATKT1P, ATKUP1, KT1, KUP1, POTASSIUM TRANSPOTER1, POTASSIUM UPTAKE TRANSPOTER1.
RA_16163	2	167	0	*AT5G52370*	28S ribosomal S34 protein. Involved in biological process, response to cold. Located in chloroplast, mitochondrion.
RA_16173	2	407	2	*AT5G59380*	Protein containing methyl-CpG-binding domain. Has sequence similarity to human MBD proteins. The mRNA is cell-to-cell mobile. Other name:MBD6.
RA_16231	10	1028	0	*AT3G45850*	P-loop containing nucleoside triphosphate hydrolases superfamily protein. Function in: microtubule motor activity, ATP binding; Involved in: microtubule-based movement.
RA_16232	5	1399	1	*AT3G45830*	Nuclear factor kappa-B-binding-like protein.
RA_19120	**3**	**1053**	**4**	***AT1G09730***	**Encodes a SUMO protease. Other names: *ASP1*, *SPF1*. Positively regulating the transition to flowering in long and short days. Along with SPF2, its activity is required for fertility as asp1/spf2 double mutants have defects in gametogenesis and embroygenesis.**
RA_20369	**2**	**367**	**1**	***AT5G45950***	**GDSL-motif esterase/acyltransferase/lipase. Expressed during flower, leave and plant embryo development stages. Expressed in flower, leaf, plant embryo, hypocotyl and shoot tissues. Participating in pollen germination, tube growth, seed germination and floral development.**
RA_20370	2	292	0	*AT5G45960*	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates. Expressed during flower and plant embryo development. Expressed in flower, plant embryo, leaf and shoot tissues.

RA_20660	2	228	0	*unknown*	Function unknown.
RA_20751	1	453	0	*AT1G10020*	Formin-like protein (DUF1005).
RA_21335	2	620	1	*AT3G17000*	Group XIV ubiquitin-conjugating enzyme that functions negative regulation of drought stress. Other names: UBC32.
RA_22657	1	473	1	*AT3G55580*	Regulator of chromosome condensation (RCC1) family protein. Other names: TCF1, TOLERANT TO CHILLING AND FREEZING1. TCF1 encodes a member of the RCC1 gene family and is required for chromatin based gene regulation of cold responsive genes in a CBF-independent manner. It is expressed in response to cold but not ABA.
RA_23070	2	352	0	*AT1G29660*	GDSL-motif esterase/acyltransferase/lipase. Enzyme group with broad substrate specificity that may catalyze acyltransfer or hydrolase reactions with lipid and non-lipid substrates. Expressed during flower, leaf development and plant embryo globular stage. Expressed in flower, leaf, plant embryo, root, shoot and stem tissues. Participating in both early and late stages of female gametophyte development[53].
RA_23103	1	136	0	*AT3G19900*	Hypothetical protein.
RA_23104	1	960	2	*AT3G22980*	Ribosomal protein S5/Elongation factor G/III/V family protein.
RA_24027	1	476	0	*AT5G63290*	Radical SAM superfamily protein. Function in oxidoreductase activity, iron-sulfur cluster binding, coproporphyrinogen oxidase activity, catalytic activity; Involved in oxidation reduction, porphyrin biosynthetic process. Located in chloroplast, cytoplasm.

The seven bolded genes are involved in flower development (Table 2).

j-sites: the number of non-introgressable sites within the gene.

L(aa): amino acid sequence length of the gene.

[1] Site: No. of highly differentiated amino acids between *R. mucronata* and *R. stylosa* are given (see also Supplementary Fig. S12).

Table S11 Summary of non-introgressable blocks (j-blocks), including singleton blocks.

No. of j-blocks (No. scaffolds with j-blocks)	1,189 (168)
Length of j-blocks - Range (mean)	10 bp – 149.4 Kb (1,365 bp)
No. of j-sites in a block – Range (total non-i sites)	1 - 7 bp (1,398 bp)
Total length of j-blocks (% of the genome)	1,622,664 bp (0.71%)
No. of genes within j-blocks	395
No. of genes containing j-sites	263

A j-block has >= 1 non-introgressable sites (j-sites).

Supplementary Figures

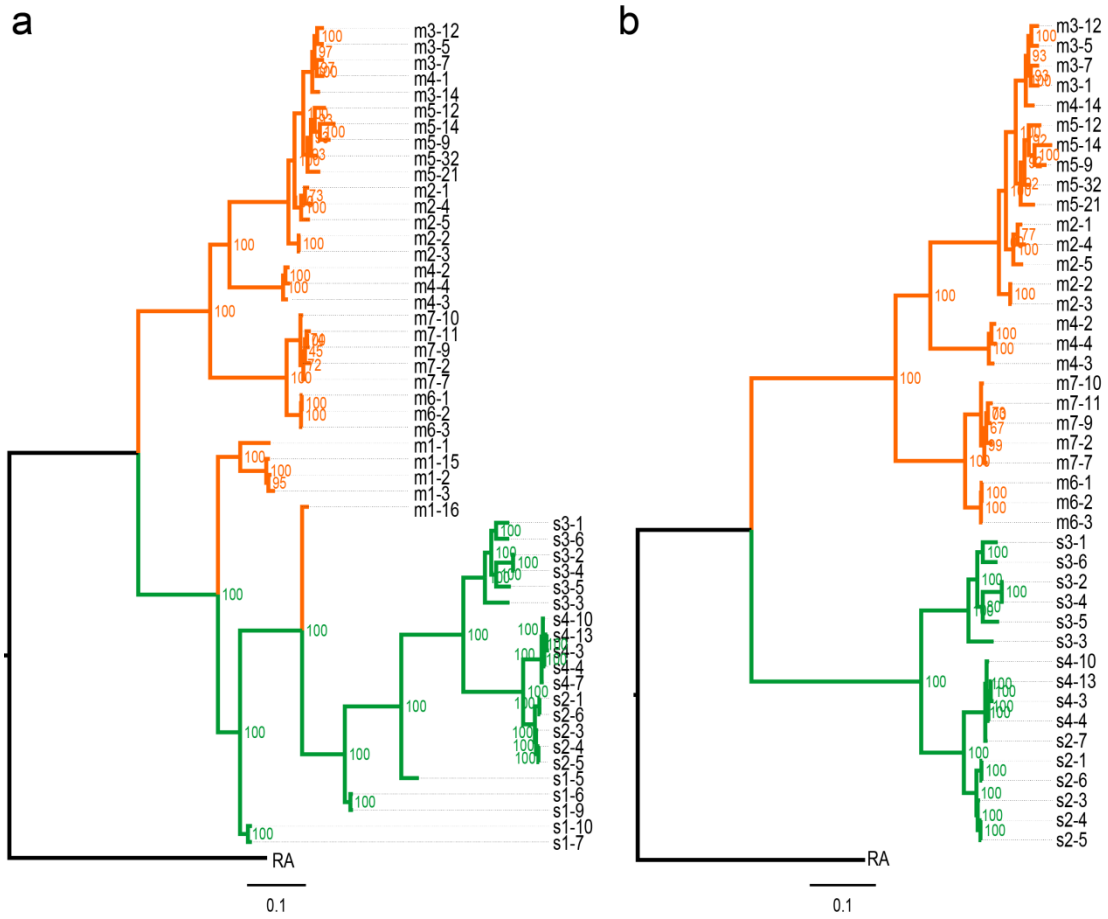

Fig. S1. Phylogenetic relationships of *R. mucronata* and *R. stylosa* samples with (a) or without (b) sympatric populations m1 and s1. Branches of *R. mucronata* populations (or individuals) are colored in orange while those of *R. stylosa* are in green. The Maximum Likelihood (ML) trees were generated by IQTREE[39]. Bootstrap values are provided on each branch.

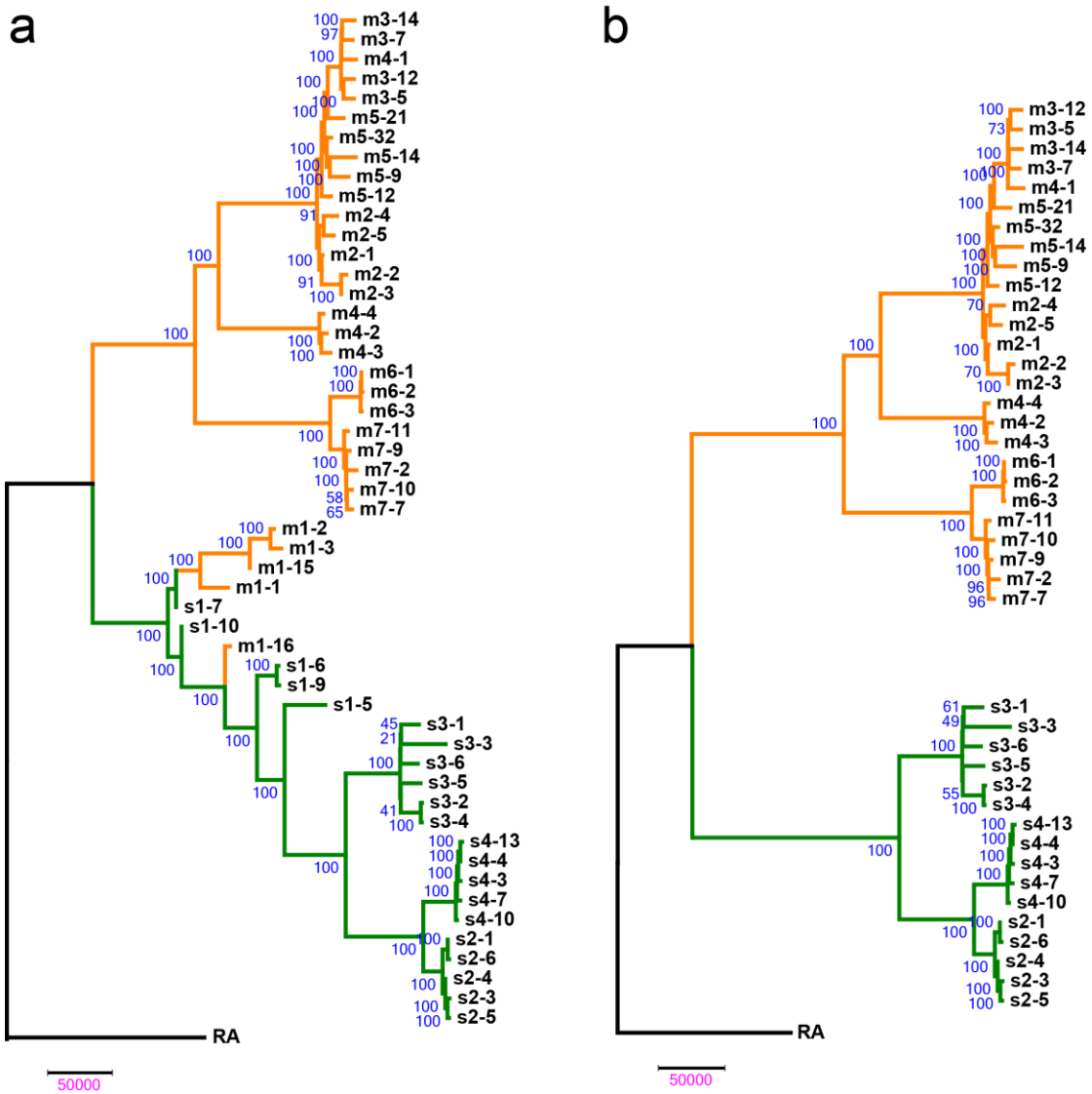

Fig. S2. Phylogenetic relationships of *R. mucronata* and *R. stylosa* samples with (a) or without (b) sympatric populations m1 and s1. Branches of *R. mucronata* populations (or individuals) are colored in orange while those of *R. stylosa* are in green. The Neighbor-joining (NJ) trees were generated by MEGA7[40]. Bootstrap values are provided on each branch.

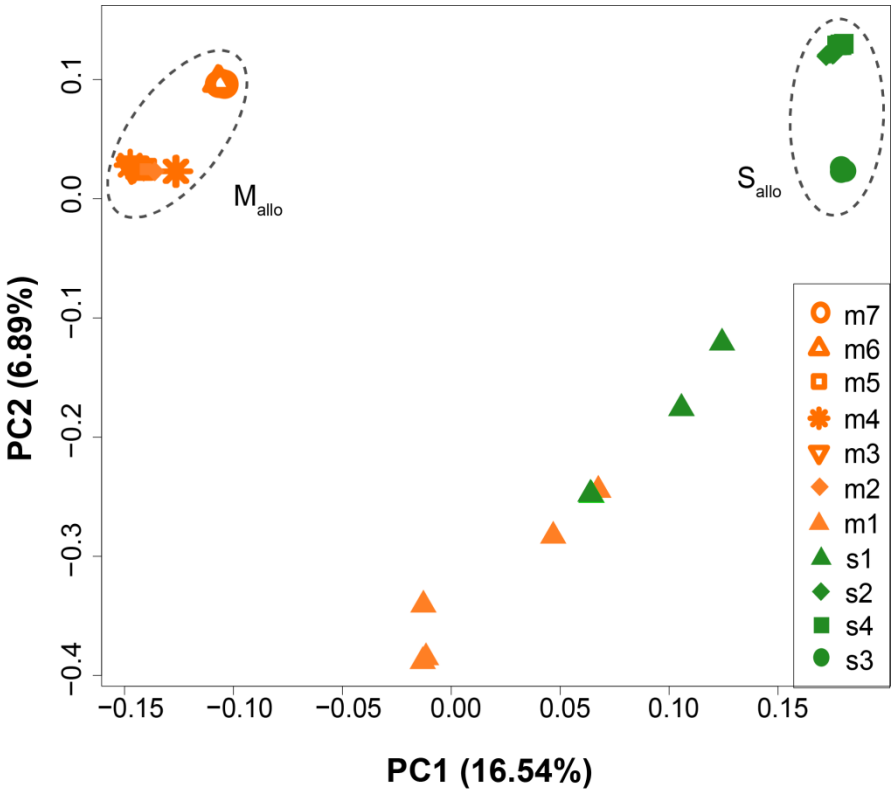

Fig. S3. PCA plot of all populations[41]. *R. mucronata* individuals are colored in orange while *R. stylosa* individuals in green. Allopatric populations (M_{allo} and S_{allo}) are highlighted by dotted lines.

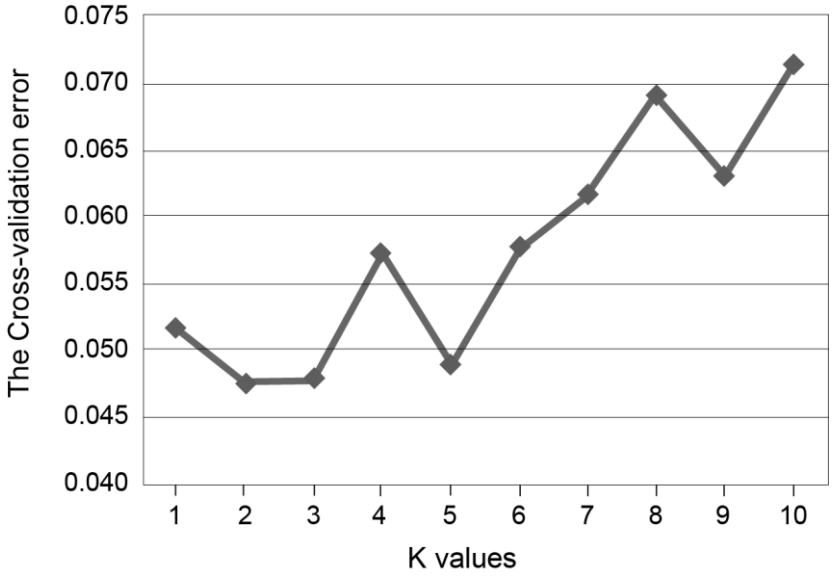

Fig. S4. Cross-validation errors corresponding to different *K* values in ADMIXTURE[44]. The best K is 2 with the lowest cross-validation score.

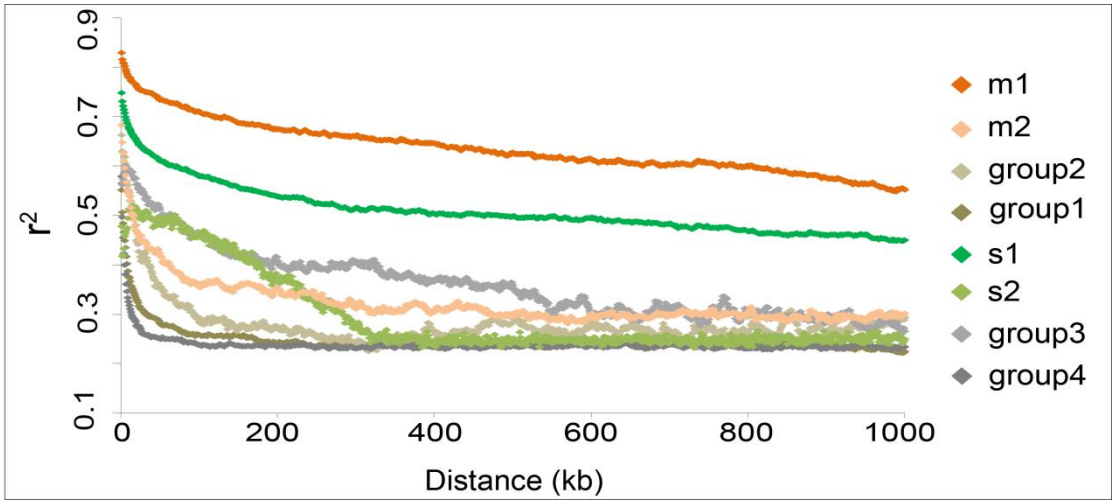

Fig. S5. Decay of linkage disequilibrium in *R. mucronata* and *R. stylosa* populations measured by r^2. Four groups: group1 contains population m6 and m7; group2 includes population m2, m3, m4, and m5; group3 represents population s2 and s4; and group4 is population s3. The populations in the same group are genetically closely related to each other.

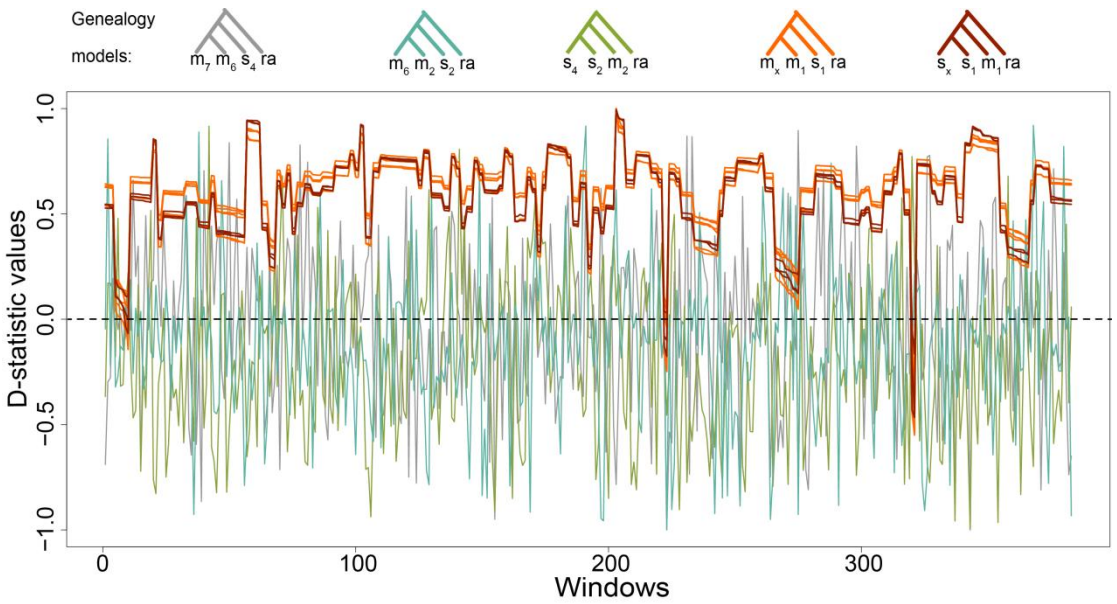

Fig. S6. Patterson's *D* statistic scan across the genome for the genealogy models above, showing genome wide evidence of gene flow between sympatric populations m1 and s1. The models colors correspond to the curves colors. In the last two genealogy models, m_x represents *R. mucronata* population m2, m3, m4, m5, m6 or m7, and s_x represents *R. stylosa* population s2, s3 or s4 (see Supplementary Table S7 for detail information).

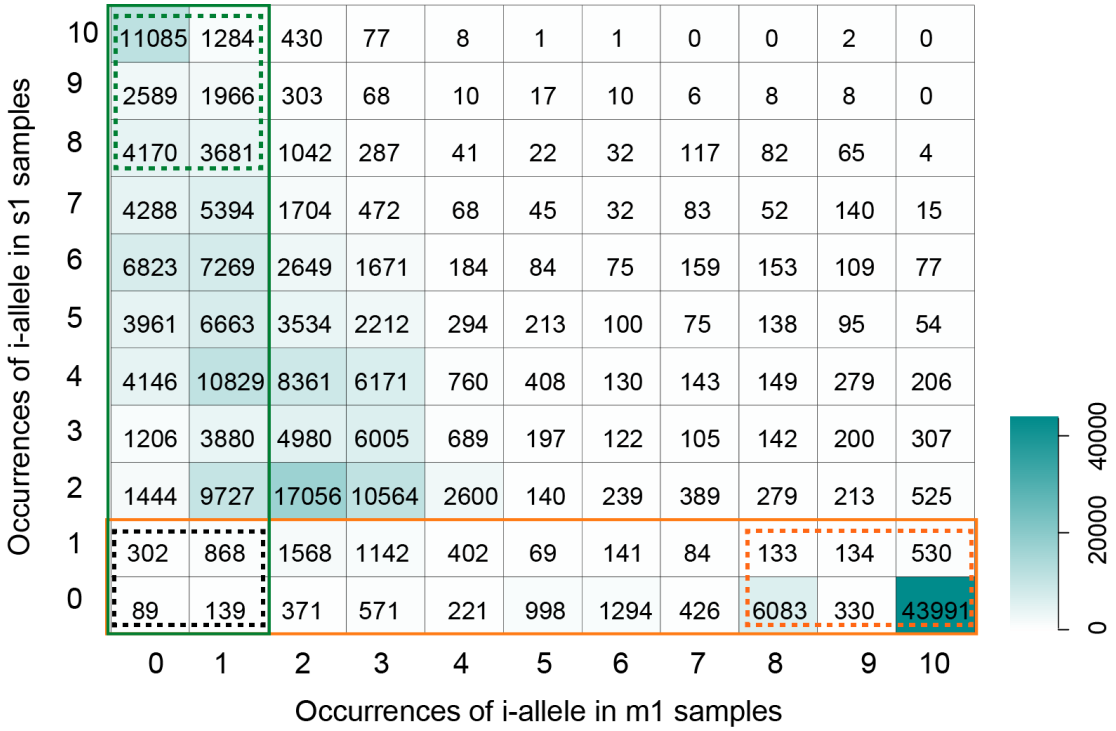

Fig. S7. The site distributions in m1 (orange) and s1 (green) samples, classified by the occurrence of the i-allele (introgressed allele), which ranges from 0 to 10 (given five diploid individuals, or 10 genomes). The actual numbers of sites are shown. Sites in the orange (in m1) and green (in s1) solid boxes correspond to the site distributions in Fig. 3B. The orange and green dotted boxes contain the i-sites (with >=8 occurrences of i-allele) in m1 and s1 populations, respectively. The black dotted box shows non-introgressable sites (j-sites) with <=1 occurrences of i-allele both in m1 and s1 samples.

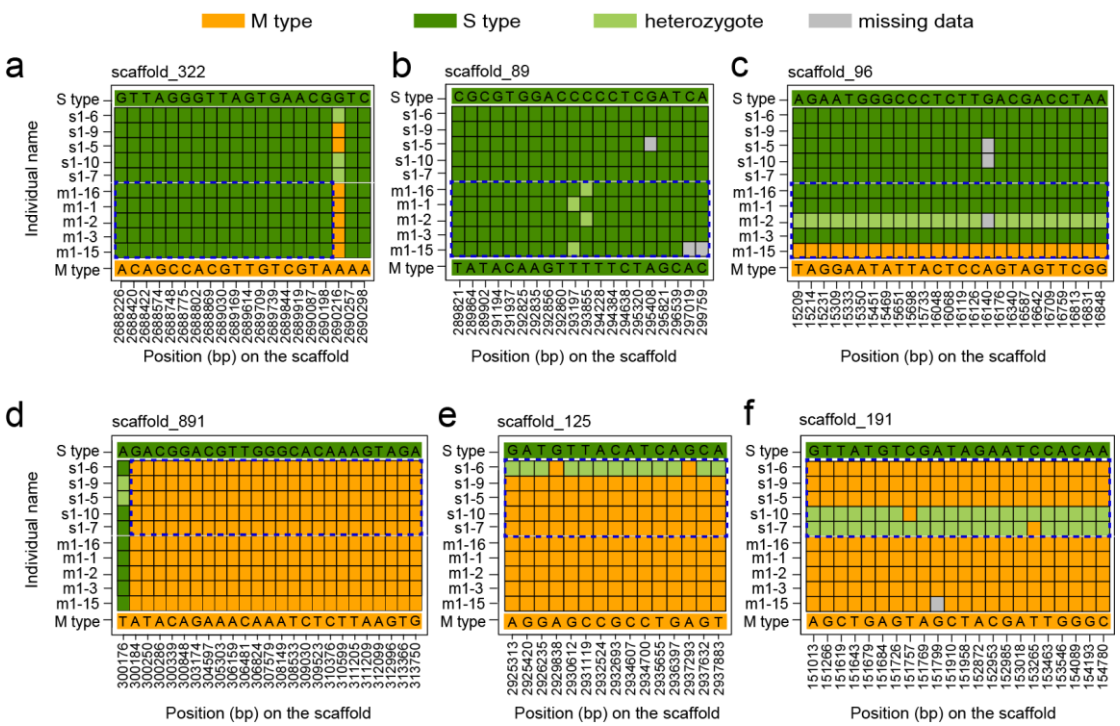

Fig. S8. Examples of i-blocks (in blue dotted boxes) in m1 genomes (a, b, and c) and in s1 genomes (d, e, and f) at the site level. Only the d- and i-sites are displayed. Each site is color-coded for the MM, MS, and SS type (M for the *R. mucronata* variant and S for the *R. stylosa* variant, see Materials and Methods).

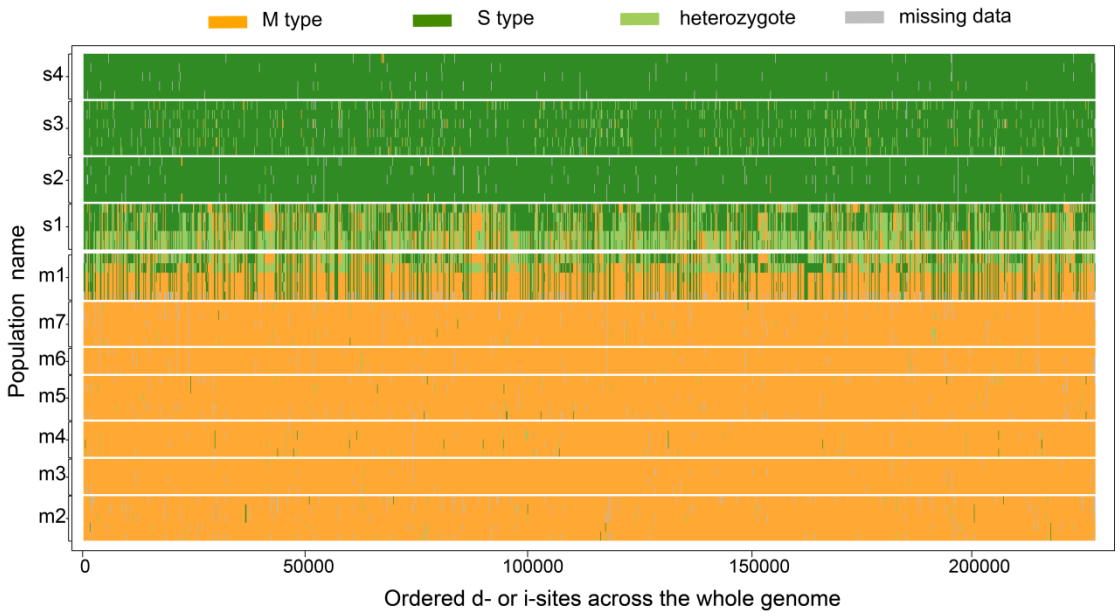

Fig. S9. The genome-wide landscape of i-blocks in m1 and s1 samples. All 52 *R. mucronata* and *R. stylosa* individuals are shown. In each ideogram, all d- and i-sites are displayed consecutively. Each site is color-coded for the MM, MS, and SS types as in Fig. S10.

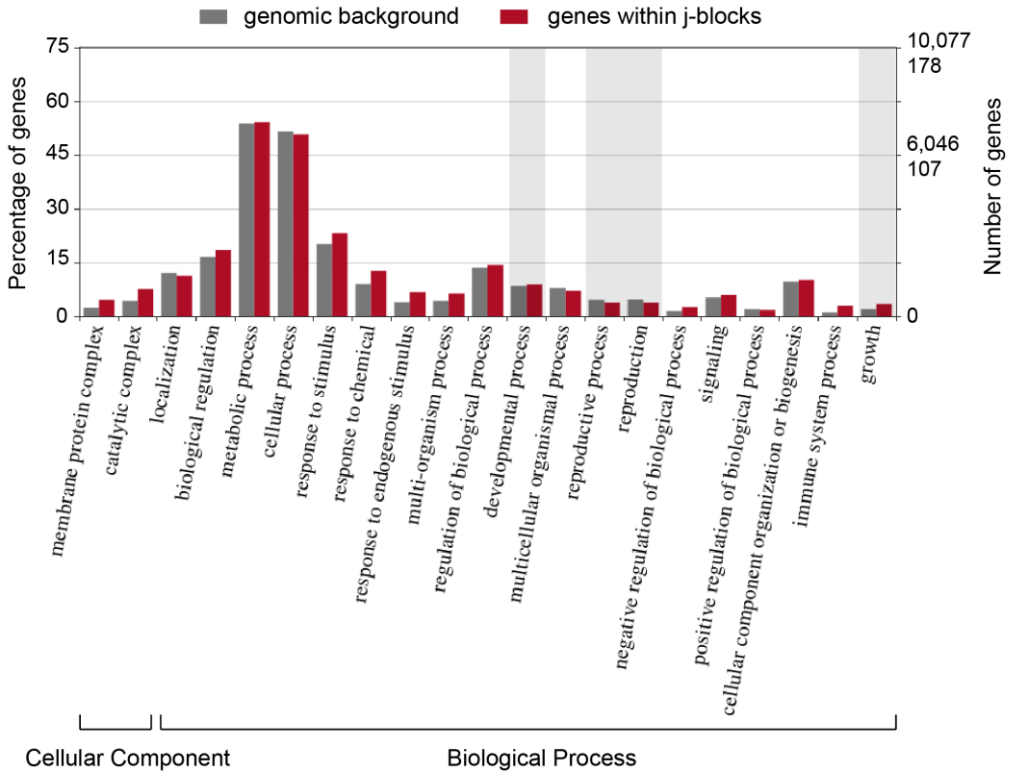

Fig. S10. GO (gene ontology) enrichment analysis of all genes in the j-blocks (or non-introgressable blocks) in Table S11, using WEGO 2.0 (Web Gene Ontology Annotation Plot, http://wego.genomics.org.cn/)[83].

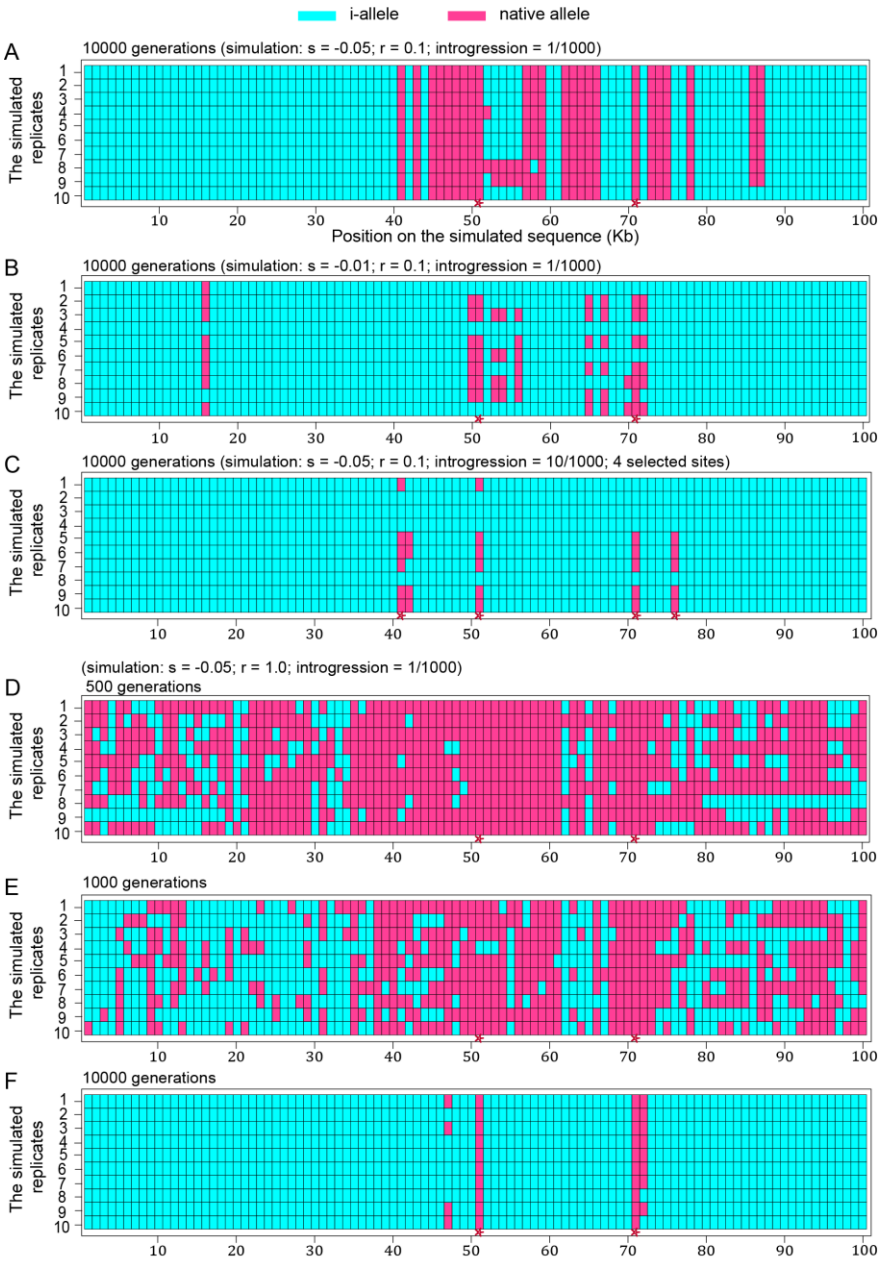

Fig. S11. Simulated introgressions in haploid 100 Kb genomes. Speciation genes are marked by red stars at the bottom. Sites of introgression and non-introgression are marked blue and pink, respectively. (A) Simulated results of 10000 generations under strong selection (s = -0.05), low recombination rate (r = 0.1 for per 100Kb per generation), and low introgression (1/1000 per generation). Native alleles are not purified at non-selected sites. (B) Simulated results of 10000 generations under weak selection (s = -0.01), low recombination rate (r = 0.1 for per 100Kb per generation), and low introgression (1/1000 per generation). Selected sites have introgressions, too. (C) Simulated results of 10000 generations under strong selection (s = -0.05) + four loci under selection (#41, #51, #71 and #76), low recombination rate (r = 0.1 for per 100Kb per generation), and high introgression (10/1000 per generation). Selected sites have introgressions as well. (D-F) Simulated results under strong selection (s = -0.05), high recombination (r = 1.0 for per 100Kb per generation), and low introgression (1/1000 per generation). Three time points are given. This is closest to the expected pattern.

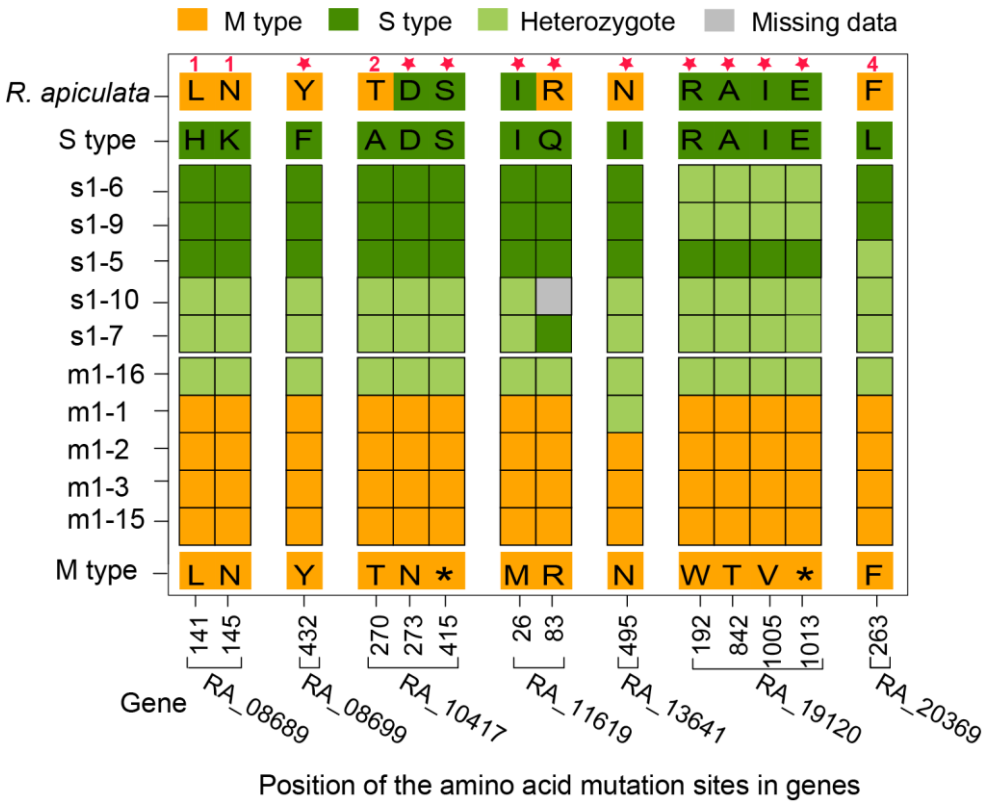

Fig. S12. Highly differentiated amino acids between *R. mucronata* and *R. stylosa* samples in the seven genes involved in flower development. Sites marked by red stars are fixed between allopatric *R. mucronata* (m2-m7) and *R. stylosa* (s2-s4) samples, and the rest each has 1-4 (number in red) heterozygotes in s3 (*R. stylosa*) population. Each site is color-coded for the M type (orange), Heterozygote (light green), and S type (green) (M for allopatric *R. mucronata* dominant amino acid and S for allopatric *R. stylosa* dominant amino acid).

Supplementary references

79. Duke, N. C. Mangrove floristics and biogeography revisited: Further deductions from biodiversity hot spots, ancestral discontinuities, and common evolutionary processes. in *Mangrove Ecosystems: A Global Biogeographic Perspective: Structure, Function, and Services* 17–53 (2017). doi:10.1007/978-3-319-62206-4_2
80. Ke, X. *et al.* Functional divergence of chloroplast Cpn60α subunits during Arabidopsis embryo development. *PLoS Genet.* **13**, (2017).
81. Hajduch, M. *et al.* Systems Analysis of Seed Filling in Arabidopsis: Using General Linear Modeling to Assess Concordance of Transcript and Protein Expression. *Plant Physiol.* **152**, 2078–2087 (2010).
82. Bryant, N., Lloyd, J., Sweeney, C., Myouga, F. & Meinke, D. Identification of Nuclear Genes Encoding Chloroplast-Localized Proteins Required for Embryo Development in Arabidopsis. *Plant Physiol.* **155**, 1678–1689 (2010).
83. Ye, J. *et al.* WEGO 2.0: A web tool for analyzing and plotting GO annotations, 2018 update. *Nucleic Acids Res.* **46**, W71–W75 (2018).