

# Toward understanding the genetic mechanisms of speciation: Fine-grained introgressions between species

**Xinfeng Wang**

Sun Yat-sen University

**Zixiao Guo**

Sun Yat-sen University

**Shaohua Xu**

Sun Yat-sen University

**Shao Shao**

Sun Yat-sen University

**Sen Li**

Sun Yat-sen University

**Ming Yang**

Sun Yat-sen University

**Qipian Chen**

Sun Yat-sen University

**Cairong Zhong**

Hainan Dongzhai Harbor National Nature Reserve Administration

**Norman Duke**

James Cook University

**Ziwen He**

Sun Yat-sen University

**Suhua Shi** (✉ [lssssh@mail.sysu.edu.cn](mailto:lssssh@mail.sysu.edu.cn))

Sun Yat-sen University

---

## Article

**Keywords:** mangroves, population genomics, speciation, introgression, species hybridization

**Posted Date:** September 10th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-71265/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Toward understanding the genetic mechanisms of speciation: Fine-grained introgressions between species

Xinfeng Wang <sup>a,\*</sup>, Zixiao Guo <sup>a,\*</sup>, Shaohua Xu <sup>a</sup>, Shao Shao <sup>a</sup>, Sen Li <sup>a</sup>, Ming Yang <sup>a</sup>, Qipian Chen <sup>a</sup>, Cairong Zhong <sup>b</sup>, Norman C. Duke <sup>c</sup>, Ziwen He <sup>a†</sup>, and Suhua Shi <sup>a†</sup>

<sup>a</sup> State Key Laboratory of Biocontrol, Guangdong Key Lab of Plant Resources, School of Life Sciences, Sun Yat-Sen University, Guangzhou, Guangdong, China

<sup>b</sup> Hainan Dongzhai Harbor National Nature Reserve Administration, Haikou, Hainan, China

<sup>c</sup> Centre for Tropical Water and Aquatic Ecosystem Research, James Cook University, Townsville, Queensland, Australia

\* These authors contributed equally to this work.

† Correspondence should be addressed to S.S. (lssssh@mail.sysu.edu.cn) or Z.H. (heziwen@mail.sysu.edu.cn)

**Abstract:** The biological species concept (BSC) is the prevailing definition of species whereby genes cannot be exchanged during or after speciation. In contrast, since only genes that contribute to the adaptive divergence between species should be non-exchangeable, BSC appears to bear little relationship to the genetic process of speciation. The rejection of BSC demands evidence of continual gene flow until the completion of speciation. Here, we carry out the sequencing and *de novo* high-quality assembly of the genomes of two closely related mangrove species (*Rhizophora mucronata* and *R. stylosa*). Whole-genome re-sequencing of individuals from their distributions on the tropical coasts shows their genomes to be well delineated in allopatry. They became sympatric in northeastern Australia where their genomes harbor 9,963 and 3,874 introgression blocks, respectively, with each block averaging only about 3-4 Kb. These fine-grained introgressions indicate that gene flow continues even after numerous loci have evolved to become non-introgressable “genomic islets”. Many of these islets, only 1.4 Kb on average, harbor “speciation loci” which contribute to the divergence in flower development or gamete production and thus result in fitness reduction upon introgression. Our fine-grained analysis may thus be the first one to show that gene flow not only happens during speciation, but often continues until the completion of speciation. Under the weight of the extensive literature, BSC may deserve to be retired.

**Keywords:** mangroves, population genomics, speciation, introgression, species hybridization

## Introduction

The biological species concept (BSC) has been the gold standard of modern evolutionary biology<sup>1–10</sup>. In this concept, species are products of allopatric speciation during which geographical isolation ensures the absence of gene flow and permits continual divergence. Here, we will use the term BSC broadly for both the process of speciation and the concept of species. BSC makes clear assumptions about the genetics of species divergence, postulating that nearly the entire genome evolves as a cohesive unit<sup>1</sup>. BSC is essentially a genomic concept of species<sup>2,5</sup>.

In an alternative genic view, species are defined by a set of loci that govern the morphological, reproductive, behavioral and ecological characters. These “speciation genes” may, collectively, account for no more than a fraction of the genome<sup>11–13</sup>. This fraction should be fitness-reducing upon introgression<sup>14</sup>, whereas the rest of the genome can be freely exchanged without a fitness consequence. In short, the diverging genomes comprise both introgressable and non-introgressable DNA segments. These non-introgressable segments are often referred to as “genomic islands” which are, in theory, more divergent than the rest of the genome<sup>2,15–18</sup>. Based on this conjecture, genomic islands have been identified in a wide array of speciation events<sup>16,17,26,27,18–25</sup>. It would seem timely to propose that, because substantial gene flow often accompanies speciation, BSC should be rejected.

In this backdrop, Wang et al. (2020)<sup>5</sup> recently offer a vigorous defense of BSC, which in fact expects gene flow during speciation because geographical isolation (for example, at the Isthmus of Panama<sup>28</sup>) could only be built up gradually. In other words, BSC only assumes that speciation should end with a period of strict isolation but does not specify how it should begin (see Fig. 1 of ref<sup>5</sup>). The extensive genomic literature is often about introgressions in the early stages of speciation, which is not incompatible with the generalized BSC. Therefore, if a definitive proof of continual gene flow throughout speciation can be obtained, the extensive literature, collectively, should be able to reject BSC. The subject has been further explored in a spirited debate on the issues raised by Wang et al.<sup>5,29–32</sup>.

The issue of “speciation with gene flow” is important at all levels of evolutionary biology, ranging from the taxonomic practice of delineating species, the mechanisms of creating biodiversity hotspots to the nature of genic interactions in speciation. We further show that, by focusing on the small genomic islands with no gene flow, many speciation genes can be efficiently identified.

## Results

Fig. 1A shows a hypothetical segment of the genome in an early, middle and late stage of speciation<sup>2,15,16,33</sup>. Importantly, by the late stages of speciation, many phenotypic characters underpinned by numerous loci of adaptive differences are assumed to have evolved. Such ensembles of functionally divergent genes between species and racial groups have been extensively reported<sup>2,11,37,38,12–14,16,33–36</sup>. In this genic view of speciation<sup>2,13</sup>, gene flow continues but the introgressions would gradually decrease in size and may become too fine-grained to identify. In this framework, we study two closely related mangrove species – *Rhizophora mucronata* vs. *R. stylosa*<sup>7,39–45</sup>. Mangroves

are woody plants that have colonized the intertidal zones of tropical coasts<sup>7,39,40,46</sup>. Because of the narrow band of suitable habitats along the coasts (or near river mouths), the global distributions of mangroves are essentially one dimensional, making them ideal for biogeographical studies of speciation.

### ***De novo assembly of the genomes of R. mucronata and R. stylosa***

We sequenced the genome of one *R. mucronata* and one *R. stylosa* individual, with estimated genome size of ~300 Mb and ~370 Mb, respectively (Supplementary Table S1; see Supplementary Table S2 for details). Although re-sequencing of these two genomes have been done before<sup>39</sup>, the aims of this study demand accurate assembly of their own genomes in order to assess possible fine-grained introgressions. We therefore carry out de-novo sequencing of their genomes and attain the chromosome-level assembly in this study. Re-sequencing of individuals from their geographical ranges of distribution is based on the new assembly.

We obtained 33.84 Gb (gigabases) and 36.13 Gb of raw data, corresponding to 142X coverage of the assembled genomes (Supplementary Table S2). For *R. mucronata*, the 237.85 Mb (megabases) of sequences is assembled into 14,496 scaffolds with N50 at 12.03 Mb and 18 scaffolds (chr1-18)  $\geq$  5 Mb. The assembled chr1-18 account for 84.38% (200.70 Mb) of the genome and corresponds to the diploid chromosome number of  $2n = 36$  (Fig. 1B, Supplementary Table S1 and Fig. S1). The assembled *R. stylosa* genome is 253.54 Mb, containing 9,750 scaffolds with N50 at 12.59 Mb. The top 18 scaffolds account for 219.66 Mb, or 86.63% of the genome (Fig. 1B, Supplementary Table S1 and Fig. S2). The estimated genome coverage and mapping rate are both high at 94.2% and 94.52% for *R. mucronata*. For *R. stylosa*, the numbers are 92.74% and 94.3% (Supplementary Tables S1 and S2). These results indicate high-quality de novo genomes approaching chromosome-level completeness. A total of 26,540 and 30,375 protein-coding genes in the *R. mucronata* and *R. stylosa* genome (Supplementary Table S1) are, respectively, classified into 19,089 and 19,018 gene families.

To see the evolution of the genomic structures of *R. mucronata* and *R. stylosa* in the larger phylogenetic context, we compare their genomes to those of *Carallia longipes* (unpublished data), *Bruguiera gymnorrhiza*<sup>46</sup> and *Rhizophora apiculata*<sup>39</sup>; *C. longipes* being from one of the closest non-mangrove genera in the Rhizophoraceae family. In total, 20,971 gene families are identified among the five species, with 6,014 are single-copy. Using these single-copy orthologs, we reconstruct the phylogeny and estimate divergence times (Fig. 1C). The phylogeny shows that *R. mucronata* and *R. stylosa* are the most closely-related pair diverging at ~3.78 Mya (with the 95% confidence interval at [3.15, 4.35] Mya; Fig. 1C and Supplementary Table S3 and Fig. S3). The inter-specific Ks (~0.0031) and the genomic divergence ( $D_{xy} = 0.0031$ ) all support the close relationship between *R. mucronata* and *R. stylosa* (Supplementary Table S2 and Fig. S4). We identify 663 collinear blocks between the two species that harbor 18,705 genes in *R. mucronata* and 18,952 genes in *R. stylosa* (Fig. 1B).

### ***Population genomic diversity within R. mucronata and R. stylosa***

*Rhizophora mucronata* has a wide distribution in the Indo-Western Pacific (IWP), particularly to the west of the Strait of Malacca and all the way to East Africa. In contrast, *R. stylosa* extends eastward from the Strait of Malacca to the western Pacific Islands (Fig. 2A). The two species have been reported to overlap in scattered locales along a number of western Pacific coastlines. However, in our own field trips, their relative abundance is often skewed in favor of one species and their co-occurrence has been rarely found. The sole exception in our collection is in the Daintree River (DR) area of northeastern Australia, where both species are quite abundant (Fig. 2A). It is the evidence from this site of sympatry that is instructive about the genetic makeup of species.

For population genomic studies, 21 *R. stylosa* individuals from four locations (labeled s1-s4) and 31 *R. mucronata* individuals from seven locations (named m1-m7) were analyzed (Fig. 2A and Supplementary Table S4). Note that m1 and s1 refer to the sympatric DR samples. Whole genomes of all samples were sequenced on the Illumina HiSeq 2000 platform, yielding a mean depth of 16X (ranging from 12X to 22X) (Supplementary Table S5 and Table S6). The short reads of each individual were mapped to the de novo *R. mucronata* genome, with average genomic coverage of 81% (80%-83%) (Supplementary Table S4). The level of genetic diversity shows two patterns. Low genetic diversity is found in all allopatric populations (average  $\theta_{\pi}$  at 0.62 and 0.60 per Kb for *R. mucronata* m2-m7 and *R. stylosa* s2-s4, respectively (see also Supplementary Tables S2 and S4). The level is much higher in the sympatric DR populations ( $\theta_{\pi} = 1.37/\text{Kb}$  and  $2.09/\text{Kb}$ , respectively). The Watterson's estimates ( $\theta_w$ ) are similar (Supplementary Table S4, see Materials and Methods).

### ***Divergence between the two species in allopatry***

The genomic divergence between the two species is  $4.14 \times 10^{-3}$  (Supplementary Table S7). We first constructed a Maximum Likelihood (ML) tree using RAXML<sup>47</sup> on the sequences of the 31 *R. mucronata* and 21 *R. stylosa* individuals from the 11 populations. The ML tree bifurcates with a clear delineation between species across all allopatric populations. However, the m1 and s1 (i.e., DR) samples show strong signs of admixture as they are “in the middle” of the bifurcated tree (Fig. 2A). When the DR samples are removed, the phylogeny shows clear delineation (Fig. 2B). These two trees are robust when rebuilt using the ML method in IQTREE<sup>48</sup> or the Neighbour-Joining (NJ) method in MEGA7<sup>49</sup> (Supplementary Fig. S5 and S6). The monophyletic delineation of *R. mucronata* and *R. stylosa* in allopatry is also supported by the principal component analysis (PCA)<sup>50</sup> (Supplementary Fig. S7).

In total, 1.7 million variable sites are detected across all populations of the two species (Supplementary Table S8). We first partition these sites by excluding the DR samples (see Materials and Methods). Each site is then represented by an  $F_{ST}$  value with  $F_{ST} = 0$  indicating no differentiation between the two species in allopatry and  $F_{ST} = 1$  indicating complete differentiation. Figure 2C shows the U-shaped distribution whereby the abundance of sites at the far right reveals the extensive

differentiation between species. Such a U-shape distribution is typical of species of some divergence with little gene flow <sup>51</sup>.

Morphologically, the two species are distinguished by style length <sup>40,41</sup>, as pictured in Fig. 3A. The morphological differences between *R. mucronata* and *R. stylosa* across populations are shown in Fig. 3C. *R. mucronata* is readily distinguished by its short style, in the range of 0.9-1.6mm (Fig. 3A). In contrast, the style of *R. stylosa* is long, 2.4-5.3 mm (Fig. 3A) with no overlap between the two species (Fig. 3C) <sup>40,41</sup>. While the style length varies from locale to locale in both species, this trait is a species-diagnostic one across locales. The two species also show different habitat preferences with *R. mucronata* usually found upriver while *R. stylosa* is close to the river mouth (Fig. 3D). Additional diagnostic morphological characters, which are less stable, are listed in the Supplement (Table S9).

### ***Characterizations of R. mucronata and R. stylosa in sympatry (the DR samples)***

Morphological characters of the sympatric DR samples, which appear admixed in their DNA sequences, remain distinct. The style length of each sample is concordant with that of the allopatric populations of the same species (Fig. 3B). In fact, the two species in all sympatric populations can be clearly delineated by this character (see Fig. 3C). In the DR area, these two species are parapatric-sympatric with distributions up- or down-river and extensive overlaps in the middle (Fig. 3D). This difference in habitat preference is seen everywhere <sup>41,42</sup>.

Corroborating the phylogenetic positions of the DR samples in Fig. 2A, we use the Bayesian clustering analysis, ADMIXTURE <sup>52</sup>. The analysis identifies two genetic components that make up the genomes of the DR samples (Fig. 4A). PCA results also indicate significant admixture in m1 and s1 individuals (Supplementary Fig. S7). Furthermore, because species divergence is monophyletic in all allopatric comparisons, incomplete lineage sorting as the cause of the observed admixture in the DR samples is rejected. In short, we interpret the high  $F_{ST}$  sites as manifesting the divergence after speciation (Fig. 2C) with subsequent admixture in the DR area. Additional tests of introgression (LD analysis,  $D$  statistic and the modified  $f_d$  statistic) are presented in the Supplement (Tables S8, S10 and Figs. S8-S9).

### ***Extensive introgressions in sympatry***

For the two species in sympatry in the DR area, we ask the following questions: 1) How many introgressed segments can be found in each species? 2) Is the introgression symmetric between the two species? 3) How fine-grained are the introgressed segments (i.e., many small segments or a few large ones)? A few large blocks are expected after recent hybridizations but many fine-grained blocks may result from old introgressions that have been eroded by recombination. If true, introgression (and hence speciation itself) might be a prolonged process. 4) How many genomic segments are non-introgressable and what are their genic contents? Question 4 will be the subject of the next section.

To quantify the introgressions in the DR area, we first define the divergent sites (or d-sites) between *R. stylosa* and *R. mucronata* in allopatry. Among the d-sites, we can then define the introgression sites (i-sites) between the m1 and s1 samples in sympatry. There are 305,418 d-sites, defined as sites with  $F_{ST} > 0.8$  between the two species in allopatry. Note that the bulk of d-sites (212,626 sites) are fully divergent with  $F_{ST} = 1.0$  (Fig. 2C). An i-site is where the introgressed allele (or i-allele) is found in  $\geq n$  of the 10 genomes in the m1 or s1 samples, where  $n$  is usually equal to 8. (Note that both m1 and s1 samples have five diploid individuals, or 10 genomes.) In general, if introgression is observed in one direction, say from *R. stylosa* to *R. mucronata* in the DR area, the same site usually does not show introgression ( $n \leq 1$ ) in the reciprocal direction from *R. mucronata* to *R. stylosa* (see Materials and Methods).

We now define  $n$  (the number of genomes carrying the introgressed allele), which is necessary for determining the i-sites. It is obviously better to set  $n$  close to the maximum of 10 for strongly penetrant introgressions. Fig. 4B shows the level of introgressions in the two directions. We set  $n = 8$  for the m1 samples where the i-allele is usually found  $\geq 8$  times (the orange bars in Fig. 4B). Hence, the results with  $n = 2$  and  $n = 8$  would not be very different. Furthermore, to avoid the confounding presence of remnant ancient polymorphisms, we require introgressions at an i-site to be strongly asymmetric:  $\geq n$  one way (say, from *R. stylosa* to *R. mucronata*) and  $\leq 1$  in the reciprocal direction (Supplementary Fig. S10). For the *R. stylosa* (s1) samples, the occurrence of the i-allele is rather even between 2 and 10 (the green bars in Fig. 4B). The asymmetry is probably due to the geography of the DR area, which is at the fringe of the *R. mucronata* distribution. Consequently, gene flow from *R. mucronata* into *R. stylosa* may be more limited here, resulting in the lower frequency of introgressions in the s1 samples. In this regard, setting  $n = 8$  would miss many introgressions in *R. stylosa* leading to a much lower introgression rate than in *R. mucronata*. Nevertheless, the final estimations appear robust even when  $n$  is set as low as 2 (see below). Simulations of these scenarios are presented in the Supplement.

Clearly, introgressions do not happen site-by-site, but appear as long segments of DNA consisting of consecutive i-sites. We shall label these segments “introgression blocks” (or i-blocks) (Supplementary Fig. S11). Fig. 5A shows a segment of the genome that comprises a string of d-sites and i-sites as defined above. These d-sites and i-sites are embedded in a background of low- $F_{ST}$  or invariant sites shown as dots in Fig. 5A. This figure shows 3 i-blocks, each consisting of one, two or three i-sites. The length of each block is defined by the distance between the two breakpoints flanking the block. Unless specified, we remove the singleton i-blocks that harbor only a single i-site when presenting the length distribution of i-blocks.

The analysis of i-blocks is summarized in Table 1 (see also Supplementary Tables S11-S13). We shall focus on the results with  $n = 8$  but the results with  $n = 2$  and  $n = 10$  are given for comparison. In the DR area, samples of *R. mucronata* (m1) harbor far more introgressions than those of *R. stylosa* (s1). The bottom of Table 1 at  $n=8$  shows that 16.09 or 23.09% of the *R. mucronata* genomes are introgressions from *R. stylosa*, the two values depending on whether singleton i-blocks are counted. In the opposite direction, 7.97-12.06% of the *R. stylosa* genomes are introgressions. The introgressions of Table 1 can be visualized in Figs. 5-6. The salient observation is the highly fine-grained nature of the

introgressions. In *R. mucronata*, the introgressions are distributed over 9,963 i-blocks with an average length of 3.24 Kb. In *R. stylosa*, there are 3,874 i-blocks with an average size of 4.13 Kb. During the evolution, there should be numerous recombination events that break the introgressions into thousands of tiny i-blocks. It should be noted that Table 1 and Figs. 5-6 present only the extreme cases of introgressions that rise to very high frequencies.

The distributions of i-blocks are shown at the large genomic scale in Fig. 5B, at the scaffold scale in Fig. 5C and as individual sites in Fig. 6A-6C. Note that only d-sites and i-sites are portrayed in these figures, which convey the visual impression of the fine-grained nature of the i-blocks. (As shown in Fig. 2C, the d- and i-sites are the 305,418 sites with  $F_{ST} > 0.8$ ; the rest are invariant and lowly divergent sites.) Specifically, the i-blocks are dispersed across the whole genome (Fig. 5B and Supplementary Fig. S12). Indeed, all top 18 scaffolds harbor the switching between i- and d-blocks both in m1 and s1 genomes, massively and frequently (Fig. 5B and Table 1). Figure 5C shows that the switching between i- and d-blocks can occur in a few to tens of Kbs. At the site level, i-blocks and d-blocks can switch within a small distance (Fig. 6A-6C). An i-block (or d-block) may harbor only one i-site (or d-site), referred to as singleton block (Fig. 6A-6C, Table 1 and Supplementary Table S11). Singleton blocks, not uncommon but less reliable, are not used in the tally.

The extensive fine-grained introgressions convey two messages. First, hybridizations may happen continually over a long span of time. Each hybridization event would initially bring in whole-chromosome introgressions that are subsequently broken down by recombination. Small DNA fragments may have been introgressed in this piece-meal manner continually. Second, loci underlying differential adaptation between species may be very common such that introgressions tend to be small, and thus free of the introgressed alleles that are deleterious in the genetic background of another species<sup>14</sup>. In the next section, we will direct the attention toward non-introgressions, which are blocks of native alleles flanked by introgressed DNA segments.

### ***Very fine-grained interspersions between “introgressable” and “non-introgressable” blocks***

Some DNA segments may not be introgressable due to the presence of genes of adaptive significance. Such loci, by definition, contribute to reproductive isolation or ecological speciation<sup>2,9</sup> and have sometimes been referred to “speciation genes”<sup>8,10,12,13,34,53,54</sup>. The number, size and direction of introgressions are therefore functions of a number of parameters: 1) the rate of hybridization; 2) the strength of selection against the speciation genes when introgressed; 3) the number and location of speciation loci; 4) the rate of recombination that free neutral genes from the linkage to speciation genes; and 5) the length of time since the time of initial hybridization.

In a companion study<sup>5</sup>, we carry out computer simulations based on the Recurrent Selection and Backcross (RSB) model<sup>55</sup> (see also Materials and Methods). The RSB model is proposed for identifying genes of complex traits<sup>55</sup>. In its execution, one dilutes the genome of breed A (say, the bull dog) with that of breed B (e.g., the border collie) but retains all the desired phenotypic traits of the former. This is done by continually selecting for the traits of breed A while backcrossing the culled

products to breed B. The scheme is almost identical with the process of “speciation with gene flow” in its model structure. They differ only in the parameter values; for example, the length of time in speciation is far greater and gene flow is much smaller, and often bidirectional. The differences necessitate separate simulations for speciation with gene flow. As shown in Fig. 2 of Ref. <sup>5</sup> and Fig. 6D-6E in this study (see also Supplementary Fig. S13 and Materials and Methods), introgressions are fine-grained around almost all speciation genes. The patterns resemble the observations reported in this study.

With the j-site defined above, a j-block is defined as a DNA segment containing at least one j-site (Table 2). By these stringent criteria, there are ~1200 j-blocks which together account for < 1% of the genome and harbor 328 coding genes of which 171 genes contain j-sites (Table 2, Supplementary Table S14 and Fig. S14). For higher confidence, we also show j-blocks with at least two j-sites (Table 2). While only 19 genes containing j-sites are found in these j-blocks (Supplementary Table S15), it is remarkable that 6 of the 19 genes function in flower development and/or gamete production and development as shown in Table 2 (see the WEGO gene ontology in Supplementary Fig. S14, where a larger set of genes is presented under less stringent criteria). One (*RM\_77078.7*) of the six genes, known as *EMF1* (embryonic flower 1), regulates reproductive development and is involved in controlling flowering development <sup>56,57</sup>. *RM\_76773.10*, *RM\_76979.9* (*NAC2*) and *RM\_77530.24* are involved in regulating the stamen development, pollen germination and tube growth <sup>58-60</sup>. *RM\_76929.10*, *RM\_76979* and *RM\_77333.68* all play a role in embryonic development <sup>61-63</sup>. Since all six genes contain highly differentiated amino acids and non-introgressable sites (j-sites) (Table 2, Supplementary Table S15 and Fig. S15), their involvement in the speciation between *R. mucronata* and *R. stylosa* seems plausible.

## Discussion

The species of *R. mucronata* and *R. stylosa* in the DR area are unusual among sympatric species reported in the literature <sup>40,44,45</sup> as explained below. These features may be the reason that they are ideal for a rigorous test of BSC by looking for fine-scale introgressions in sympatry, vis-à-vis the populations in allopatry (Fig. 2A and Fig. 5B). Most important, the fine-scale analysis permits the identification of many genomic islets (~1.4 kb on average) that harbor speciation loci. Genic divergence at these loci contributes to the divergence in flower development or gamete production.

The DR populations stand out even among populations from other sympatric locations between the two species. For example, the m2/s2 collections from Singapore both show the expected phylogenetic relationship of their species designation (Fig. 2A and Fig. 2B). This pattern of little admixture is consistently found in other locales: in Brandan, Indonesia <sup>44</sup>, in Panay Island, Philippines, in Kosrae, Micronesia, in Yap, Micronesia and in North Sulawesi, Indonesia <sup>45</sup>. The two species in sympatry outside of the DR area occasionally show a slight tendency of being “on the fringe” of their phylogenetic cluster, thus suggesting low-level introgressions. Nevertheless, the extensive fine-grained introgression observed in the DR samples has not been reported before. Importantly, Yan

et al. <sup>45</sup> did notice samples from northeastern Australia (from Trinity Inlet and Daintree River, Australia) to be different without further clarification.

The near absence of prior reports of fine-grained introgression is understandable as several conditions have to be met for this phenomenon to be realized. The first condition is that the two diverging populations have to be in the right stage when they first come into contact. This stage roughly corresponds to Stage III defined in Wu (2001) <sup>2</sup> whereby speciation is nearly complete but gene flow is still possible. Had the secondary contact happened before this stage (in Stage II, for example), the process of speciation could be arrested or reversed. On the other hand, if the contact starts too late, there would be too little gene flow to have the extensive introgressions observed in the DR area.

Among the sympatric locations reported for *R. mucronata* and *R. stylosa*, northern Australia has been suggested as the place where the two species first came into contact in their incipient stage of speciation (see the Supplement) <sup>40</sup>. In this interpretation, the two diverging taxa moved eastward either along the northern coasts of Indian Ocean to Southeast Asia or by crossing the Indian Ocean to Australia <sup>40</sup>. *R. mucronata* in Southeast Asia then dispersed south and eastward to Australia, while *R. stylosa* in Australia migrated further east and north into SE Asia (see the Supplement) <sup>40</sup>. By the time the two diverging species were reunited in sympatry outside of northern Australia, the two species would have developed some reproductive isolation which hindered exchanging of genes, thus explaining the unmixed phylogenetic relationship in these other locales.

The second condition may be even more difficult to satisfy – that the two species need to remain in secondary contact for a long period of time <sup>4</sup>. As discussed, numerous recombination events accumulated over a long period of time are necessary to achieve the fine-grained introgression. Non-trivial gene flow during secondary contact may also prevent further build-ups of functional divergence that would lead to the complete cessation of gene flow <sup>2</sup>.

The third condition is ecological. Two sympatric species without niche separation would face the problem of competitive exclusion <sup>64</sup>, making long-term coexistence unlikely. *R. mucronata* and *R. stylosa* had evolved a degree of niche separation that results in incomplete overlaps in habit preference (Fig. 3D). Given the necessary confluence of all these conditions, *R. mucronata* and *R. stylosa* in the DR area may be truly exceptional. Instead of a few large “genomic islands” <sup>19–21,24–26</sup>, we observe in the DR samples a large number of tiny islets. A precedent is the genomic island surrounding the speciation gene, *Odysseus*, which is < 2 Kb in size <sup>16</sup> and, hence, an islet. Small islets are obviously conducive for the identification of “speciation genes” as only a few candidate genes are involved (see Table 2 and Supplementary Table S15).

These conditions make this study particularly suited to the inquiry of speciation with continual gene flow. It raises the question of whether the extensive reports on speciation with gene flow <sup>15,16,27,65–70,19–26</sup> should be taken as evidence against BSC. As Wang et al.’s survey <sup>5</sup> shows, many of these studies are likely about gene flow in the early stages. The rest is not conclusive as to when the

inferred gene flow actually happened. Together with the fine-grained pattern shown in this study, the extensive literature<sup>15,16,27,65–70,19–26</sup> may indeed offer sufficient empirical evidence against BSC.

Finally, BSC's requirement of complete cessation of gene flow, imposed by strict geographical isolation, has posed many dilemmas. There are simply too few stable geographical features like the Isthmus of Panama to facilitate strict allopatric speciation<sup>6,7</sup>. Furthermore, such a geographical feature does not help to explain the teeming biodiversity on the Indo-western Pacific coasts<sup>7,71</sup>, in the Amazons<sup>6</sup> or in Lake Victoria<sup>72</sup>. In this sense, BSC is conceptually flawed as pointed out repeatedly<sup>2</sup>. Furthermore, genomic data that have sufficient resolution to provide a meaningful test have consistently failed to support BSC<sup>7,13,14,73–77</sup>. Had BSC been an imperfect guide in the actual practice of species delineation, it may still be useful, but it is not. It may therefore be time to seriously consider abandoning BSC in the curricula of evolutionary biology.

**Acknowledgements:** We thank Loren Rieseberg, James Mallet, Matthew Hahn, Richard Abbott, Patrik Nosil, Jeffrey Feder, Trevor Price, Nicholas Barton and Nolan Kane for insightful comments. We thank Wei Lun Ng for the photo of *R. mucronata* style in Fig. 2A. This study was supported by the National Natural Science Foundation of China (91731301, 31600182 and 31830005); the National Key Research and Development Plan (2017FY100705); Guangdong Basic and Applied Basic Research Foundation (2019A1515010752); the China Postdoctoral Science Foundation (2019M663212) and the Chang Hungta Science Foundation of Sun Yat-Sen University.

**Data Availability:** Genomic raw reads of *R. mucronata* and *R. stylosa* individuals have been deposited to the Genome Sequence Archive of National Genomics Data Center & BIG Data Center with GSA accession number CRA001688 (<http://bigd.big.ac.cn/gsa>).

## References

1. Mayr, E. *Animal Species and Evolution*. (Cambridge, MA: Harvard University Press, 1963).
2. Wu, C.-I. The genic view of the process of speciation. *J. Evol. Biol.* **14**, 851–865 (2001).
3. Futuyma, D. J. & Kirkpatrick, M. *Evolution*. (Sunderland, MA: Sinauer Associates, Inc., 2017).
4. Coyne, J. A. & Orr, H. A. *Speciation*. (Sunderland, MA: Sinauer Associates, Inc., 2004).
5. Wang, X., He, Z., Shi, S. & Wu, C.-I. Genes and speciation – Is it time to abandon the Biological Species Concept? *Natl. Sci. Rev.* **7**, 1387–1397 (2020).
6. Haffer, J. Speciation in amazonian forest birds. *Science* **165**, 131–137 (1969).
7. He, Z. *et al.* Speciation with gene flow via cycles of isolation and migration: Insights from multiple mangrove taxa. *Natl. Sci. Rev.* **6**, 275–288 (2019).

8. Presgraves, D. C. The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* **11**, 175–180 (2010).
9. Schluter, D. Evidence for ecological speciation and its alternative. *Science* **323**, 737–741 (2009).
10. Mallet, J. What does Drosophila genetics tell us about speciation? *Trends Ecol. Evol.* **21**, 386–393 (2006).
11. Wu, C. & Palopoli, M. F. Genetics of Postmating Reproductive Isolation in Animals. *Annu. Rev. Genet.* **28**, 283–308 (2003).
12. Ting, C. T., Tsaur, S. C., Wu, M. L. & Wu, C.-I. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282**, 1501–1504 (1998).
13. Wu, C. & Ting, C. T. Genes and speciation. *Nat. Rev. Genet.* **5**, 114–122 (2004).
14. Fang, S. *et al.* Incompatibility and competitive exclusion of genomic segments between sibling Drosophila species. *PLoS Genet.* **8**, e1002795 (2012).
15. Osada, N. & Wu, C.-I. Inferring the Mode of Speciation From Genomic Data. *Genetics* **169**, 259–264 (2005).
16. Ting, C.-T., Tsaur, S.-C. & Wu, C.-I. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5313–5316 (2002).
17. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
18. Yang, M., He, Z., Shi, S. & Wu, C.-I. Can genomic data alone tell us whether speciation happened with gene flow? *Mol. Ecol.* **26**, 2845–2849 (2017).
19. Ellegren, H. *et al.* The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756–760 (2012).
20. Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biol.* **3**, 1572–1578 (2005).
21. Harr, B. Genomic islands of differentiation between house mouse subspecies. *Genome Res.* **16**, 730–737 (2006).
22. Renaut, S. *et al.* Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat. Commun.* **4**, 1827 (2013).
23. Clarkson, C. S. *et al.* Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nat. Commun.* **5**, 4248 (2014).
24. Toews, D. P. L. *et al.* Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Curr. Biol.* **26**, 2313–2318 (2016).

25. Poelstra, J. W. *et al.* The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**, 1410–1414 (2014).
26. Malinsky, M. *et al.* Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**, 1493–1498 (2015).
27. Carneiro, M. *et al.* The genomic architecture of population divergence between subspecies of the European Rabbit. *PLoS Genet.* **10**, e1003519 (2014).
28. O’Dea, A. *et al.* Formation of the Isthmus of Panama. *Sci. Adv.* **2**, e1600883 (2016).
29. Mallet, J. Alternative views of biological species: reproductively isolated units or genotypic clusters? *Natl. Sci. Rev.* **7**, 1401–1407 (2020).
30. Gao, L. & Rieseberg, L. H. While neither universally applicable nor practical operationally, the biological species concept continues to offer a useful framework for studying speciation. *Natl. Sci. Rev.* **7**, 1390–1400 (2020).
31. Butlin, R. K. & Stankowski, S. Is it time to abandon the biological species concept? No. *Natl. Sci. Rev.* **7**, 1400–1401 (2020).
32. Wu, C.-I., Wang, X., He, Z. & Shi, S. Replies to the commentaries on the question of “Is it time to abandon the biological species concept?” *Natl. Sci. Rev.* **7**, 1407–1409 (2020).
33. Sawamura, K., Davis, A. W. & Wu, C.-I. Genetic analysis of speciation by means of introgression into *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 2652–2655 (2002).
34. Sun, S., Ting, C. T. & Wu, C.-I. The normal function of a speciation gene, *Odysseus*, and its hybrid sterility effect. *Science* **305**, 81–83 (2004).
35. Küpper, C. *et al.* A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2015).
36. Coulthart, M. B. & Singh, R. S. High level of divergence of male-reproductive-tract proteins, between *Drosophila melanogaster* and its sibling species, *D. simulans*. *Mol. Biol. Evol.* **5**, 182–191 (1988).
37. Hollocher, H., Ting, C. T., Wu, M. L. & Wu, C.-I. Incipient speciation by sexual isolation in *Drosophila melanogaster*: Extensive genetic divergence without reinforcement. *Genetics* **147**, 1191–1201 (1997).
38. Tao, Y., Zeng, Z. B., Li, J., Hartl, D. L. & Laurie, C. C. Genetic dissection of hybrid incompatibilities between *Drosophila simulans* and *D. mauritiana*. II. Mapping hybrid male sterility loci on the third chromosome. *Genetics* **164**, 1399–1418 (2003).
39. Xu, S. *et al.* The origin, diversification and adaptation of a major mangrove clade (Rhizophoreae) revealed by whole-genome sequencing. *Natl. Sci. Rev.* **4**, 721–734 (2017).

40. Duke, N. C., Lo, E. & Sun, M. Global distribution and genetic discontinuities of mangroves - Emerging patterns in the evolution of *Rhizophora*. *Trees - Struct. Funct.* **16**, 65–79 (2002).
41. Duke, N. C. Indo-West Pacific stilt mangroves: *Rhizophora apiculata*, *R. mucronata*, *R. stylosa*, *R. x annamalai*, *R. x lamarckii*. In: Elevitch CR (ed) *Traditional trees of Pacific Islands: their culture, environment, and use* Pp.641-660 (Permanent Agriculture Resources, Holualoa, Hawaii, 2006).
42. Duke, N. C., Ball, M. C. & Ellison, J. C. Factors influencing biodiversity and distributional gradients in mangroves. *Glob. Ecol. Biogeogr. Lett.* **7**, 27–47 (1998).
43. Duke, N. C. ‘*World Mangrove iD: expert information at your fingertips*’ App Store Version 1.2, July 2017. *MangroveWatch Publication, Australia - e-book.* (2017).
44. Wee, A. K. S. *et al.* Genetic differentiation and phylogeography of partially sympatric species complex *Rhizophora mucronata* Lam. and *R. stylosa* Griff. using SSR markers Phylogenetics and phylogeography. *BMC Evol. Biol.* **15**, 1–13 (2015).
45. Yan, Y.-B., Duke, N. C. & Sun, M. Comparative Analysis of the Pattern of Population Genetic Diversity in Three Indo-West Pacific *Rhizophora* Mangrove Species. *Front. Plant Sci.* **7**, 1–17 (2016).
46. He, Z. *et al.* Convergent adaptation of the genomes of woody plants at the land-sea interface. *Natl. Sci. Rev.* (2020). doi:10.1093/nsr/nwaa027
47. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
48. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
49. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**, 1870–4 (2016).
50. Galinsky, K. J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am. J. Hum. Genet.* **98**, 456–472 (2016).
51. Futuyma, D. J. *Evolution.* (Sunderland, MA: Sinauer Associates, Inc., 2005).
52. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–64 (2009).
53. Nosil, P. & Schluter, D. The genes underlying the process of speciation. *Trends Ecol. Evol.* **26**, 160–167 (2011).
54. Wittbrodt, J. *et al.* Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu locus in *Xiphophorus*. *Nature* **341**, 415–421 (1989).

55. Luo, Z. W., Wu, C.-I & Kearsley, M. J. Precision and high-resolution mapping of quantitative trait loci by use of recurrent selection, backcross or intercross schemes. *Genetics* **161**, 915–929 (2002).
56. Sánchez, R., Kim, M. Y., Calonje, M., Moon, Y. H. & Sung, Z. R. Temporal and spatial requirement of EMF1 activity for arabidopsis vegetative and reproductive development. *Mol. Plant* **2**, 643–653 (2009).
57. Park, H. Y. *et al.* EMF1 interacts with EIP1, EIP6 or EIP9 involved in the regulation of flowering time in Arabidopsis. *Plant Cell Physiol.* **52**, 1376–1388 (2011).
58. Wang, Y. *et al.* Transcriptome Analyses Show Changes in Gene Expression to Accompany Pollen Germination and Tube Growth in Arabidopsis. *Plant Physiol.* **148**, 1201–1211 (2008).
59. Mandaokar, A. *et al.* Transcriptional regulators of stamen development in Arabidopsis identified by transcriptional profiling. *Plant J.* **46**, 984–1008 (2006).
60. Hsiao, A. S., Yeung, E. C., Ye, Z. W. & Chye, M. L. The Arabidopsis cytosolic Acyl-CoA-binding proteins play combinatory roles in pollen development. *Plant Cell Physiol.* **56**, 322–333 (2015).
61. Ke, X. *et al.* Functional divergence of chloroplast Cpn60 $\alpha$  subunits during Arabidopsis embryo development. *PLoS Genet.* **13**, e1007036 (2017).
62. Kunieda, T. *et al.* NAC family proteins NARS1/NAC2 and NARS2/NAM in the outer integument regulate embryogenesis in arabidopsis. *Plant Cell* **20**, 2631–2642 (2008).
63. 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).
64. Hardin, G. The competitive exclusion principle. *Science* **131**, 1292–1297 (1960).
65. Martin, S. H. *et al.* Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* **23**, 1817–1828 (2013).
66. Fontaine, M. C. *et al.* Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* **347**, 1258524 (2015).
67. Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G. & Sweigart, A. L. Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genet.* **10**, e1004410 (2014).
68. Lamichhaney, S. *et al.* Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371–375 (2015).
69. Schumer, M. *et al.* Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science* **360**, 656–660 (2018).

70. Edelman, N. B. *et al.* Genomic architecture and introgression shape a butterfly radiation. *Science* **366**, 594–599 (2019).
71. Hoeksema, B. W. Delineation of the Indo-Malayan Centre of Maximum Marine Biodiversity: The Coral Triangle. In: Renema W (ed) *Biogeography, time, and place: distributions, barriers, and islands* Pp.117-178 (Springer, Dordrecht, 2007).
72. Seehausen, O. Patterns in fish radiation are compatible with Pleistocene desiccation of Lake Victoria and 14 600 year history for its cichlid species flock. *Proc. R. Soc. B Biol. Sci.* **269**, 491–497 (2002).
73. Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
74. Rieseberg, L. H. A new model of speciation. *Natl. Sci. Rev.* **6**, 289–290 (2019).
75. Abbott, R. J. A mixing–isolation–mixing model of speciation can potentially explain hotspots of species diversity. *Natl. Sci. Rev.* **6**, 290–291 (2019).
76. Price, T. Allo-parapatric speciation goes offshore. *Natl. Sci. Rev.* **6**, 289–289 (2019).
77. Barton, N. H. Is speciation driven by cycles of mixing and isolation? *Natl. Sci. Rev.* **6**, 291–292 (2019).

## Materials and Methods

### *Plant material, genome sequencing and assembly*

The sequenced *R. mucronata* and *R. stylosa* individuals were collected from Dongzhai Harbor, Hainan, China (110°35'5.79" E, 19°56'39.67" N) (Supplementary Table S2), although the *R. mucronata* individual was originally introduced from Australia<sup>78</sup>. Genomic DNA extraction from leaves was done following the CTAB method<sup>79</sup>. Total RNA was extracted from leaves, flowers and fruits using the modified CTAB method<sup>80</sup>. Short-read libraries were constructed and sequenced using the BGISEQ-500 platform. 50 Kb long-read libraries were prepared using the 10X Genomics (Illumina HiSeq X Ten) platform. Before assembling, low-quality reads, adaptor sequence, N and polyA contamination were filtered out. The 10X Genomics data was used to assemble a draft genome using SUPERNOVA<sup>81</sup>. To calibrate and refine the assembly, we used clean short reads and Hi-C reads based on the 3D-DNA (3D DNA de novo genome assembly) pipeline (<https://github.com/theaidenlab/3d-dna>). After manual check and calibration with Juicabox (<https://github.com/aidenlab/Juicebox>), we anchored 18 pseudo-chromosomes (chr1-18) both for *R. mucronata* (200.70 Mb, or 84.38%) and *R. stylosa* (219.66 Mb, or 86.63%) genomes, corresponding to their diploid chromosome number of  $2n = 36$  (Supplementary Table S1)<sup>82,83</sup>. The transcriptome data was used to perform protein-coding gene annotation (26,540 genes in *R. mucronata* and 30,375 genes in *R. stylosa*; Supplementary Table S1). Genome completeness (94.2% for *R. mucronata* and 94.3% for *R. stylosa*) was assessed using the lineage database eudicotyledons\_odb10 and the BUSCO software (<https://github.com/c-omics/busco>) (Supplementary Table S1). The raw reads were mapped to the de novo genome using the Burrows-Wheeler Aligner (BWA)<sup>84,85</sup>. The mapping rates were as high as 94.52% for *R. mucronata* and 92.74% for *R. stylosa* (Supplementary Table S2).

### *Collinearity analysis*

The collinearity analysis were done within and between *R. mucronata* and *R. stylosa* genomes. Alignment was performed on protein sequences of *R. mucronata* and *R. stylosa* using BLASTP (with a cutoff e-value of  $10^{-10}$ , identity  $\geq 40\%$ ). We then used MCSanX<sup>86</sup> to identify syntenic or collinear blocks, with blocks having at least five paired homologous genes accepted. Genomic distribution of collinear blocks was visualized using the Circos (v0.65) software (Fig. 1B and Supplementary Figs. S1-S2)<sup>87</sup>. 267 and 305 syntenic blocks were identified within the *R. mucronata* and *R. stylosa* genome, occupying up to 32.01% (8,496 of all 26,540 genes) and 29.61% (8,995 of all 30,375 genes) genes of each genome (Supplementary Figs. S1-S2). Through collinearity analysis between *R. mucronata* and *R. stylosa*, we found 663 inter-specific collinear blocks with 18,705 *R. mucronata* genes (70.48% of all 26,540 genes) and 18,952 *R. stylosa* genes (62.39% of all 30,375 genes) involved (Fig. 1B).

### ***Gene family analysis and divergence time estimation***

To estimate divergence times, we used the de novo genomes of *Carallia longipes* (unpublished data), *Bruguiera gymnorrhiza*<sup>46</sup>, *Rhizophora apiculata*<sup>39</sup>, *R. mucronata* and *R. stylosa*. We used OrthoFinder-2.2.7<sup>88</sup> to identify gene families<sup>85,89</sup>. All proteins from these five species were merged to perform an all-to-all alignment using DIAMOND (with a cutoff e-value of  $10^{-30}$ )<sup>90</sup>. 20,971 gene families are identified among the five species. 6,014 of them contain only one gene in each species (i.e., single-copy orthologs). Genes from the five species that fall into shared single-copy orthologs were aligned using MUSCLE<sup>91</sup> and codon sequences were obtained using PAL2NAL.v14<sup>92</sup>. We used JMODELTEST2<sup>93</sup> to select an appropriate nucleotide-substitution model for reconstructing the phylogeny. To reconstruct the phylogenetic tree, we used RAXML<sup>47</sup> and IQTREE<sup>48</sup> with the best-fit model (GTR+G) and 1000 bootstrap replicates (Supplementary Fig. S3). Finally, the program MCMCTREE from the PAML4.8 package<sup>94</sup> was employed to estimate divergence times (Fig. 1C, Supplementary Table S3 and Fig. S3), using "seq like (usedata = 1)", "JC69 (model = 0, alpha = 0)" and "independent rates (clock = 2)". The time constraints were based on the earliest known fossil records of mangrove lineages (hypocotyl fossils of *Bruguiera* in early Eocene (constraint 1: 47.8-56 Myr ago) and the oldest records of *Rhizophora* in the upper Eocene (constraint 2: 33.9-38.0 Myr ago)<sup>39,95-97</sup>.

### ***Sampling and genome re-sequencing***

To make the samples of *R. mucronata* and *R. stylosa* more representative, we collected individuals both in allopatry and sympatry in the Indo-West Pacific region (Fig. 2 and Supplementary Tables S4). We re-sequenced 31 *R. mucronata* individuals from seven populations and 21 *R. stylosa* individuals from four populations (Fig. 2 and Supplementary Tables S4-S6). To tell apart the two species by morphology, we observed the style length and shape in the bud and took photos (Fig. 3 and Supplementary Tables S9). Fresh leaves were sampled from individual trees and dried with silica gel. Genomic DNA extraction was done following the CTAB method<sup>79</sup>. Short-read libraries were sequenced using the Illumina HiSeq 2000 platform with insert size of 350bp and constructed following the TruSeq DNA Sample Preparation Guide. We obtained high quality sequence data for each individual genome, with coverage in the 12 to 22X range (Supplementary Tables S5-S6).

### ***SNP calling and genetic diversity detection***

To identify variants, clean reads from all 52 individuals were mapped to the de novo *R. mucronata* genome using the Burrows-Wheeler Aligner (BWA)<sup>84</sup>. SAMtools<sup>98</sup> were used to import, sort, and pair bam files and remove duplications. To obtain high quality variants, single nucleotide polymorphisms (SNPs) were called and filtered using the Genome Analysis Toolkit (GATK)<sup>99</sup> and the SAMtools/bcftools<sup>98</sup> pipeline. Only consensus SNPs called by both pipelines were retained for downstream analysis. To remove low-quality variants, we eliminated all loci that had base quality (Q) or mapping quality (q) smaller than 20. We additionally performed the following stringent filtering: 1)

at least two reads had to support the minor allele to call a heterozygote; 2) only homozygous SNPs with read depth  $\geq 2$  were retained. After filtering, we selected these high-quality sites for further analyses, with multi-allelic ( $\geq 3$ ) sites, insertions, and deletions excluded. To estimate genetic diversity in each population, we calculated  $\theta_w$  (Watterson's  $\theta_w$ ) and  $\theta_\pi$  (Nei and Li's  $\theta_\pi$ )<sup>100,101</sup> within each population (Supplementary Table S4). To estimate genomic divergence between *R. mucronata* and *R. stylosa* populations, we calculated the genetic differentiation coefficient ( $F_{ST}$ ) (Fig. 2D)<sup>100,102</sup>.

### **Detecting gene flow**

We applied Patterson's  $D$  statistic and a modified  $f_d$  statistic to quantify gene flow<sup>103,104</sup>. A positive  $D$  or  $f_d$  value is an indicator of introgression (Supplement Fig. S8 and Table S10). The basic model has three ingroups ( $P_1$ ,  $P_2$ , and  $P_3$ ) and the outgroup ( $O$ ) in the genealogical relationship  $((P_1, P_2), P_3), O$ . In our analysis,  $P_1$  and  $P_2$  are different populations from the same species *R. mucronata* (or *R. stylosa*), while  $P_3$  corresponds to the other species. The outgroup is *R. apiculata*<sup>39</sup>. Positive  $D$  values imply that  $P_2$  and  $P_3$  have more shared alleles than  $P_1$  and  $P_3$  (see Supplement Table S10 and Fig. S9). The plink-1.07<sup>105</sup> software package was used to estimate linkage disequilibrium (LD), represented by the  $r^2$  statistic within each population or group (Supplement Fig. S8). LD decay was used to test for the presence of admixture events. We also calculated LD decay in sympatric populations in Singapore (s2 and m2) and allopatric *R. mucronata* and *R. stylosa* populations as controls (Supplement Fig. S8).

### **Genomic scan for introgressed and non-introgressable blocks**

We used four predefined taxa: m1 (*R. mucronata* population in Daintree River), s1 (*R. stylosa* population in Daintree River),  $M_{\text{allo}}$  (allopatric *R. mucronata* populations m2-m7), and  $S_{\text{allo}}$  (allopatric *R. stylosa* populations s2-s4). To get a more informative data set, we filtered sites with too many missing genotypes in each taxon or low divergence ( $F_{ST} \leq 0.8$ ) between  $M_{\text{allo}}$  and  $S_{\text{allo}}$ . We retained 305,418 SNPs ( $F_{ST} > 0.8$ ) which we call divergent sites or d-sites between  $M_{\text{allo}}$  and  $S_{\text{allo}}$  (Fig. 2C). 212,626 of the d-sites are fixed ( $F_{ST} = 1.0$  and  $D_{xy} = 1.0$ ) between  $M_{\text{allo}}$  and  $S_{\text{allo}}$  (Fig. 2C). There are four possible states of each d-site: homozygous *R. mucronata* variant (MM), homozygous *R. stylosa* variant (SS), heterozygote (MS), or missing data.

We then looked for introgressed sites (i-sites) and non-introgressable sites (j-sites) among all the d-sites across m1 and s1 genomes. We have five diploid individuals (or 10 haploid genomes) from the m1 and s1 populations. We have defined allele classes as follows. **Introgressed allele (i-allele)**: *R. stylosa* variant in m1 populations or *R. mucronata* variant in s1 populations. **i-site**: an i-site in m1 or in s1 genomes is defined as  $\geq 8$  occurrences of i-allele out of the 10 genomes (Fig. 4B and Supplementary Fig. S10). **j-site**: a d-site with  $\leq 1$  occurrences of i-allele in both m1 and s1 populations (Fig. 4B and Supplementary Fig. S10). **i-block**: A genomic block in one species is considered to be introgressed from the other species if one or more i-sites continuously (without disruption by other d-sites) are present (Fig. 5A, Fig. 6A-C and Supplementary Fig. S11). The length

of an i-block is determined by the midpoint between the flanking (d-sites, i-sites) intervals (as shown in Fig. 5A). **j-block**: a genomic block with one or more j-sites continuously. We define the boundaries the same as for i-blocks.

### *Simulations of genomic sequences under hybridization, selection and recombination*

To probe the influences of hybridization, selection and recombination on genomic sequences, we carry out computer simulations based on the Recurrent Selection and Backcross (RSB) model<sup>55</sup>. We set high and low levels for each parameter. Population size was set at 1000. The length of simulated sequences is 100 Kb (for convenience, 1 Kb is the basic unit that cannot be separated by recombination). The original allele in the sequence and an i-allele from the other species are differentially labeled. Hence, at the beginning of the simulations, the sequences of all individuals are in original alleles states (100 x). After several generations of hybridization, selection and recombination, the sequences become shuffled (Fig. 6D-E and Fig. S13).

We first set a low hybridization rate (introgression or migration rate,  $m = 0.001$  per generation) and recombination ( $10E-6$  per generation between adjacent base pairs). For every generation, 999 individuals are picked from the original population and one is from the other population (or species). The recombination probability ( $r$ ) for a 100 Kb sequence is about 0.1. Since population size is 1000, there will be an average of 100 individuals with recombination in each generation. Two loci under negative selection (#51 and #71) are defined in the simulated sequences. If one or both loci harbor an i-allele, the relative fitness of this sequence is 0.95 (Fig. S13A) or 0.99 (Fig. S13B). We also examined a high introgression rate regime (10/1000). In this case, four loci (#41, #51, #71 and #76) were negatively selected (relative fitness = 0.95 for an i-allele) (Fig. S13C). The scenarios in Fig. S13A show that a lower recombination rate ( $r = 0.1$ ) increases the size of non-introgressed DNA segments, because the neutral genes near positions 51 and 71 were selected against along with the speciation loci. Figure S13B-C shows that a reduced selection intensity ( $s = -0.01$ ) or a 10-fold higher introgression rate give rise to extensive introgressions. Interestingly, partial introgressions were detected even at positions 51 and 71, where selection acts against the invading alleles.

Finally, we simulated genomic sequences under a high recombination rate ( $10E-5$ ,  $r = 1.0$  for a 100 Kb simulated sequence per generation) and a low introgression rate (1/1000 per generation). Two loci (#51 and #71) were negatively selected (relative fitness = 0.95 for an i-allele) (Fig. 6D-E and Supplementary Fig. S14D-F). The simulations suggest that, given the right parameter values (Fig. 6D-E and Supplementary Fig. S14D-F), the pattern of introgression would follow exactly the prediction based solely on selection, whereby only the alleles of the speciation loci cannot be introgressed. The rest of the genome, even right next to the speciation loci, is freely shared between species.

78. Liao, B. *et al.* Preliminary Report on Introduction of Several Alien Mangrove Plants in Dongzhai Harbour of Hainan Province. *J. Cent. SOUTH For. Univ.* **26**, 63–67 (2006).

79. Doyle, J. J. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15 (1987).
80. Yang, G., Zhou, R., Tang, T. & Shi, S. Simple and efficient isolation of high-quality total RNA from *Hibiscus tiliaceus*, a mangrove associate and its relatives. *Prep. Biochem. Biotechnol.* **38**, 257–264 (2008).
81. Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M. & Jaffe, D. B. Direct determination of diploid genome sequences. *Genome Res.* **27**, 757–767 (2017).
82. Tyagi, A. P. Cytogenetics and reproductive biology of mangroves in Rhizophoraceae. *Aust. J. Bot.* **50**, 601–605 (2002).
83. Subramanian, D. Cytological Studies of some Mangroove Flora of Tamilnadu. *Cytologia (Tokyo)*. **53**, 87–92 (1988).
84. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Prepr. arXiv* (2013). doi:arXiv:1303.3997 [q-bio.GN]
85. Wang, G. D. *et al.* Structural variation during dog domestication: Insights from gray wolf and dhole genomes. *Natl. Sci. Rev.* **6**, 110–122 (2019).
86. Wang, Y. *et al.* MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
87. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
88. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
89. Lin, T. *et al.* Genome analysis of *Taraxacum kok-saghyz* Rodin provides new insights into rubber biosynthesis. *Natl. Sci. Rev.* **5**, 78–87 (2018).
90. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59–60 (2015).
91. Edgar, R. C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
92. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
93. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. JModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **9**, 772 (2012).
94. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

95. Graham, A. Paleobotanical Evidence and Molecular Data in Reconstructing the Historical Phytogeography of Rhizophoraceae 1 . *Ann. Missouri Bot. Gard.* **93**, 325–334 (2006).
96. Collinson, M. E. *Fossil Plants of the London Clay. Palaeontological Association Field Guide to Fossils* (Palaeontological Association Field Guide to Fossils, 1983).
97. Muller, J. Fossil pollen records of extant angiosperms. *Bot. Rev.* **47**, 1–142 (1981).
98. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
99. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–303 (2010).
100. Nei, M. Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. U. S. A.* **70**, 3321–3323 (1973).
101. Nei, M. & Li, W. H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 5269–5273 (1979).
102. Wright, S. The genetic structure of populations. *Ann. Eugenetics* **16**, 97–159 (1951).
103. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
104. Green, R. E. *et al.* A draft sequence of the neandertal genome. *Science* **328**, 710–722 (2010).
105. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–75 (2007).

**Tables:**

**Table 1. Summary of high-penetrance introgressed i-blocks between sympatric species**

	>=2 occurrences of i-allele		>=8 occurrences of i-allele		=10 occurrences of i-allele	
	m1	s1	m1	s1	m1	s1
<b>No. of i-blocks</b>	9,742	8,228	9,963	3,874	9,741	2,219
<b>(No. scaffolds with i-blocks)</b>	(18)	(18)	(18)	(18)	(18)	(18)
<b>Length of i-block (mean)</b>	5 bp-944.95 Kb (3.14 Kb)	8 bp-938.88 Kb (3.49 Kb)	5 bp-944.95 Kb (3.24 Kb)	8 bp-938.88 Kb (4.13 Kb)	5 bp-944.95 Kb (3.02 Kb)	8 bp-137.48 Kb (4.15 Kb)
<b>No. of i-sites in a block (total number)</b>	2-356 bp (40,780)	2-116 bp (42,862)	2-349 bp (42,572)	2-116 bp (19,958)	2-132 bp (40,727)	2-100 bp (14,868)
<b>Total length of i-blocks (% of the genome)</b>	30.63 Mb (15.26%)	28.73Mb (14.31%)	32.29 Mb (16.09%)	16.00Mb (7.97%)	29.54 Mb (14.72%)	9.22Mb (4.59%)
<b>Total length of i-blocks<sup>a</sup> (% of the genome)</b>	45.33 Mb (22.58%)	39.61 Mb (19.74%)	46.33 Mb (23.09%)	24.21 Mb (12.06%)	43.14 Mb (21.49%)	15.57 Mb (7.76%)

Note that the species origin of introgressed alleles (i-alleles) is first defined in the allopatric populations. Hence, i-alleles in the DR area can be identified even when they are bi-directional. All i-alleles in this table are uni-directional with, for example, >= 8 in one direction, while the reciprocal direction has <= 1 i-allele. An introgressed block (i-block), unless explicitly stated, has >= 2 introgressed sites (i-sites).

The 18 scaffolds collectively account for 84.38% (200.70 Mb) of the whole genome (237.85 Mb).

<sup>a</sup>These include the singleton i-blocks.

**Table 2. High-confidence non-introgressable j-blocks for the identification of genes involved in speciation**

	>=1 j-sites	>=2 j-sites
No. of j-blocks (No. scaffolds with j-blocks)	1,189 (184)	168 (44)
Length of j-blocks - Range (mean)	3 bp-43.51 Kb (1,010 bp)	23 bp-35.75 Kb (1,062 bp)
No. of j-sites in a block – Range (total j-sites)	1-6 bp (1,443 bp)	2-6 bp (422 bp)
Total length of j-blocks (% of the genome)	1,201,823 bp (0.51%)	178,520 bp (0.075%)
No. of genes with j-sites	171	19
No. of genes of flower development with j-sites	--	6 (see below)

Gene name	j-sites	L(aa)	Site <sup>a</sup>	Function in <i>Arabidopsis thaliana</i>
<b>RM_76773.10</b> (AT2G14110)	3	255	4	Haloacid dehalogenase-like hydrolase (HAD) superfamily protein. Participating in pollen germination and tube growth <sup>58</sup> .
<b>RM_76929.10</b> (AT1G55490)	2	294	3	CPN60B. Mutants in this gene develops lesions on its leaves and accelerated cell death to heat shock stress. Participating in embryo and seed development <sup>61</sup> .
<b>RM_76979.9</b> (AT3G15510)	2	199	1	NAC2 (NAC domain containing protein 2). Involved in the regulation of stamen development and embryonic development <sup>59,62</sup> .
<b>RM_77078.7</b> (AT5G11530)	3	1,415	1	EMF1 (embryonic flower 1). Involved in regulating reproductive development <sup>56,57</sup> .
<b>RM_77333.68</b> (AT1G08520)	2	755	1	ALB1. Encoding the CHLD subunit of the Mg-chelatase enzyme which involved in chlorophyll biosynthesis. Participating in embryo and seed development <sup>63</sup> .
<b>RM_77530.24</b> (AT3G05420)	2	671	2	ACBP4 (acyl-CoA binding protein 4). Expressed and function in floral lipid metabolism. Playing combinatory roles in pollen development <sup>60</sup> .

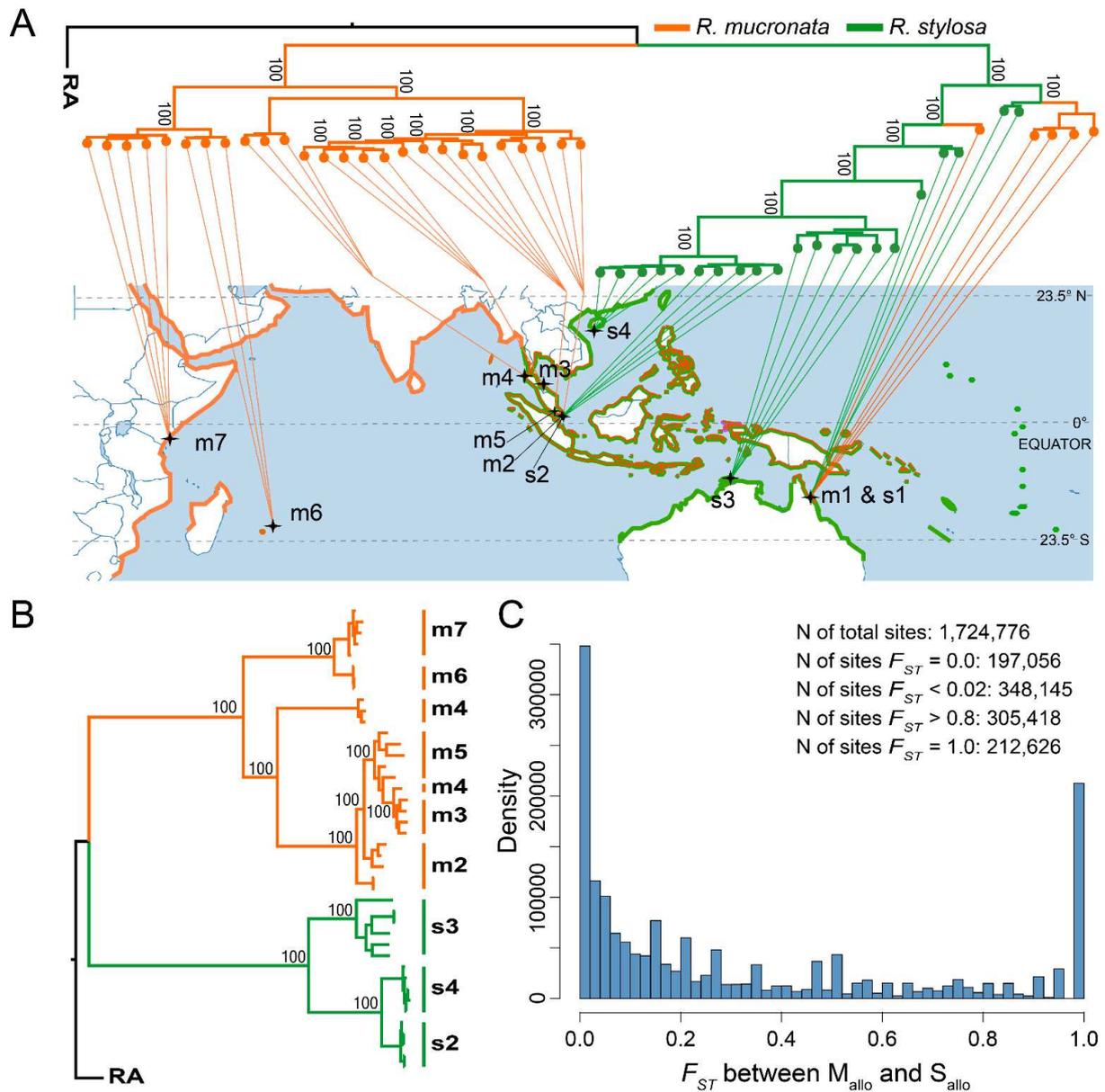
A j-block, unless explicitly stated, has >= 2 non-introgressable sites (j-sites).

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

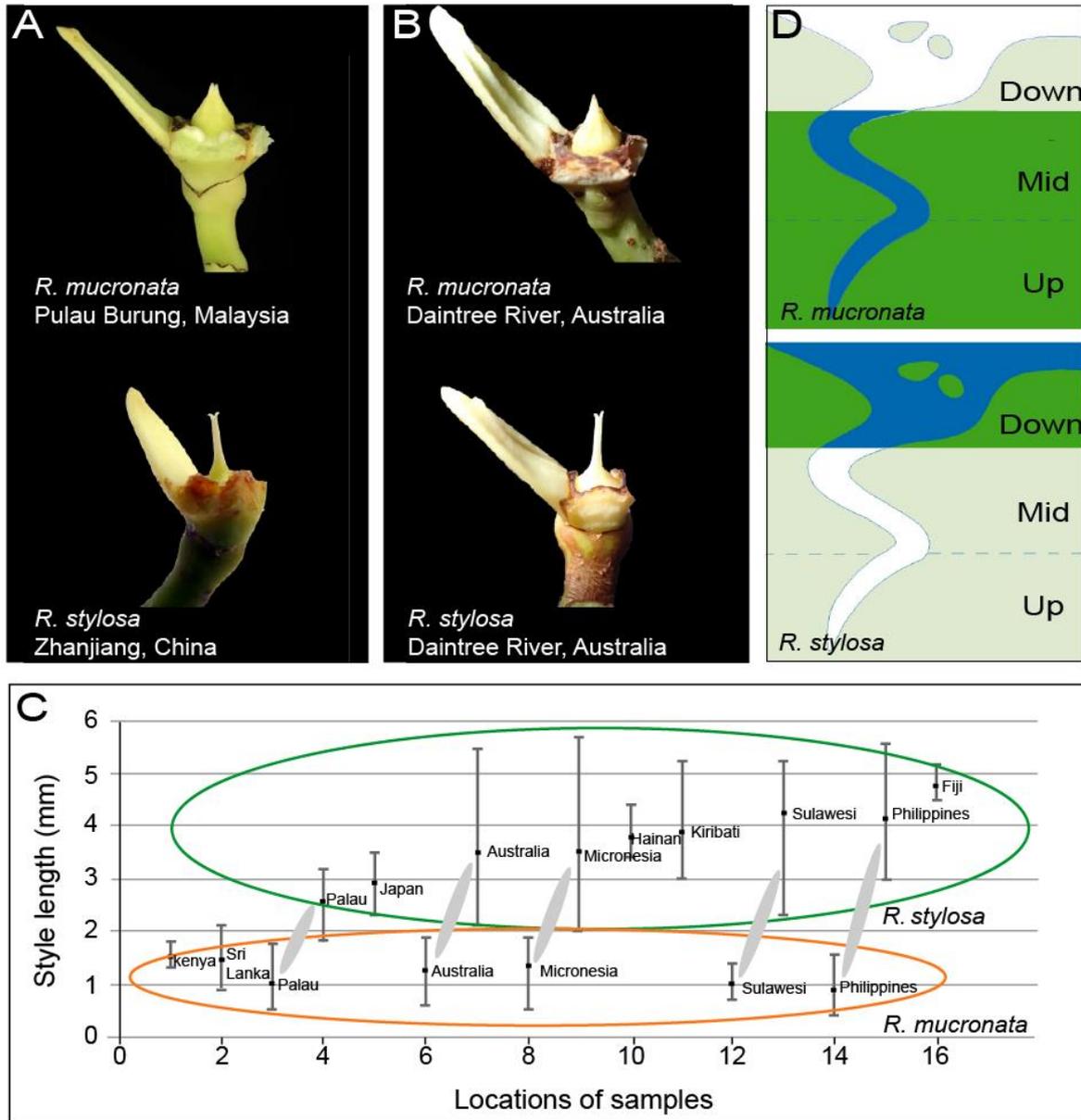
L(aa): amino acid sequence length of the gene.

<sup>a</sup> Site: No. of highly differentiated amino acids between *R. mucronata* and *R. stylosa* are given (see Supplementary Fig. S15).

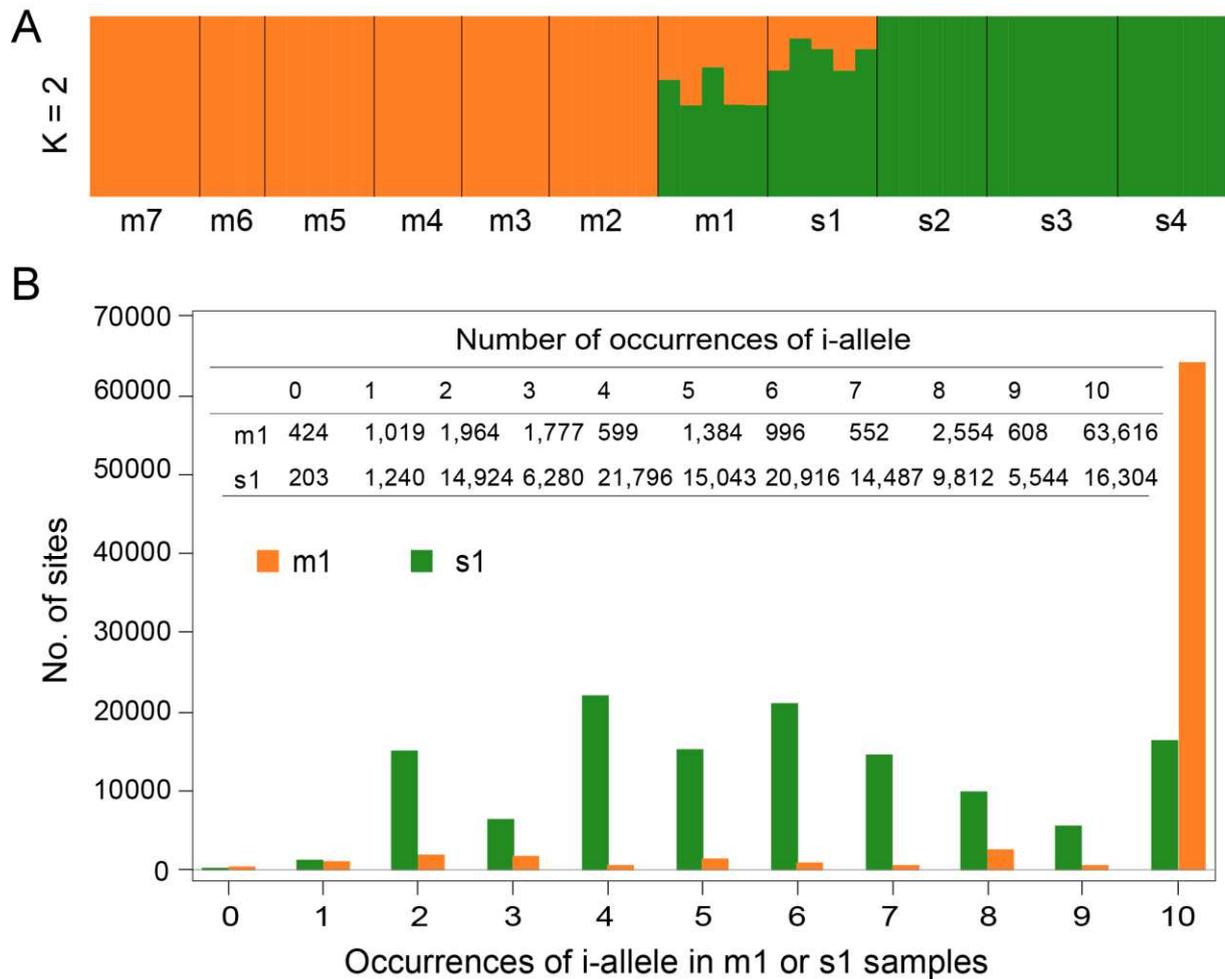




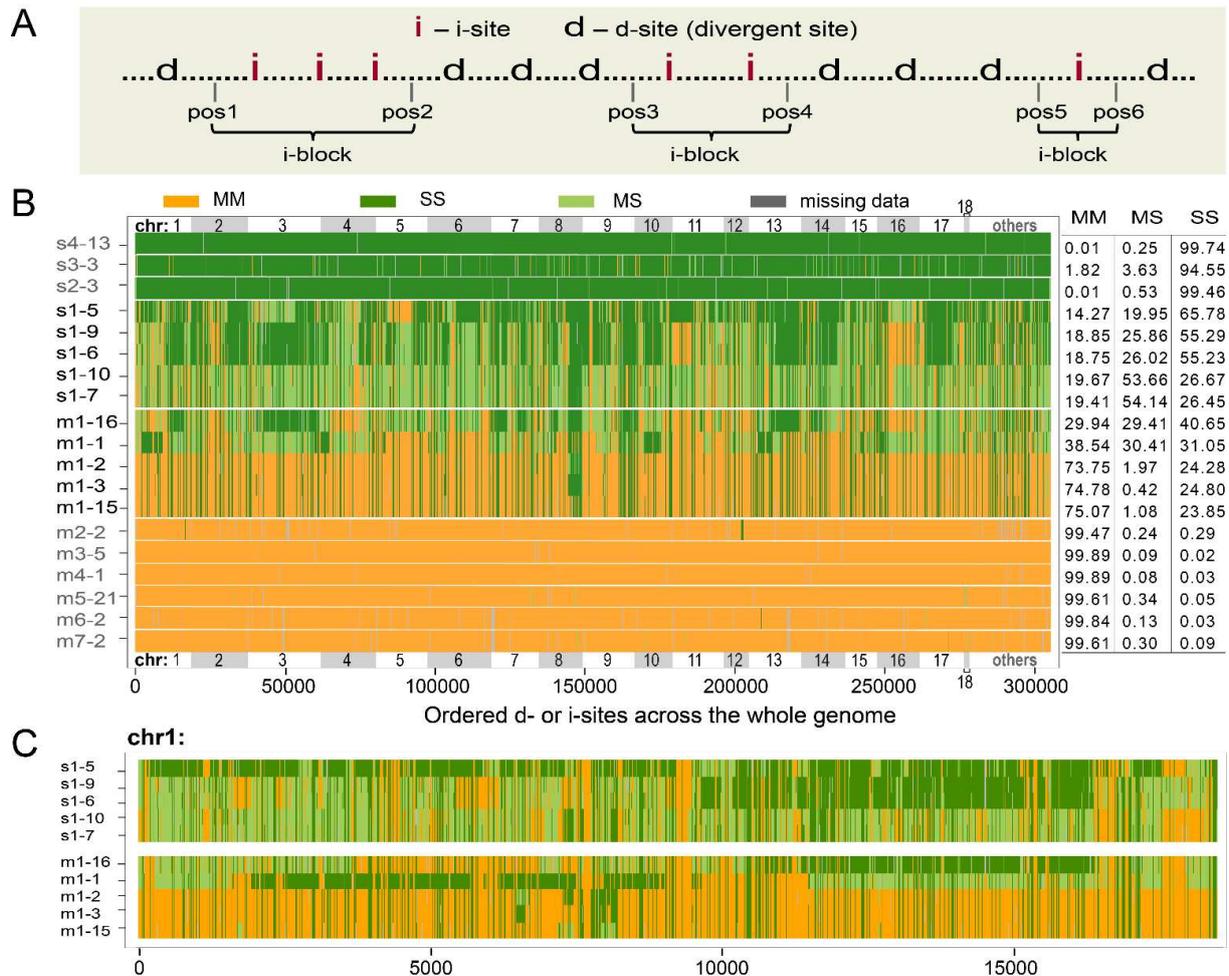
**Figure 2. Phylogenetic relationships and biogeographical observations.** (A) The biogeography and genealogy of *R. mucronata* and *R. stylosa*. The maximum-likelihood (ML) tree, generated by RAxML with 100 bootstraps, is superimposed on the biogeography. The numbers on the nodes indicate the supporting values. *R. mucronata* is colored in orange while *R. stylosa* is in green. (B) The same phylogeny excluding the sympatric m1 and s1 samples from the Daintree area. We denote the allopatric populations as  $M_{allo}$  (m2-m7) and  $S_{allo}$  (s2-s4). (C) The spectrum of the  $F_{ST}$  statistic between the  $M_{allo}$  and  $S_{allo}$  samples.



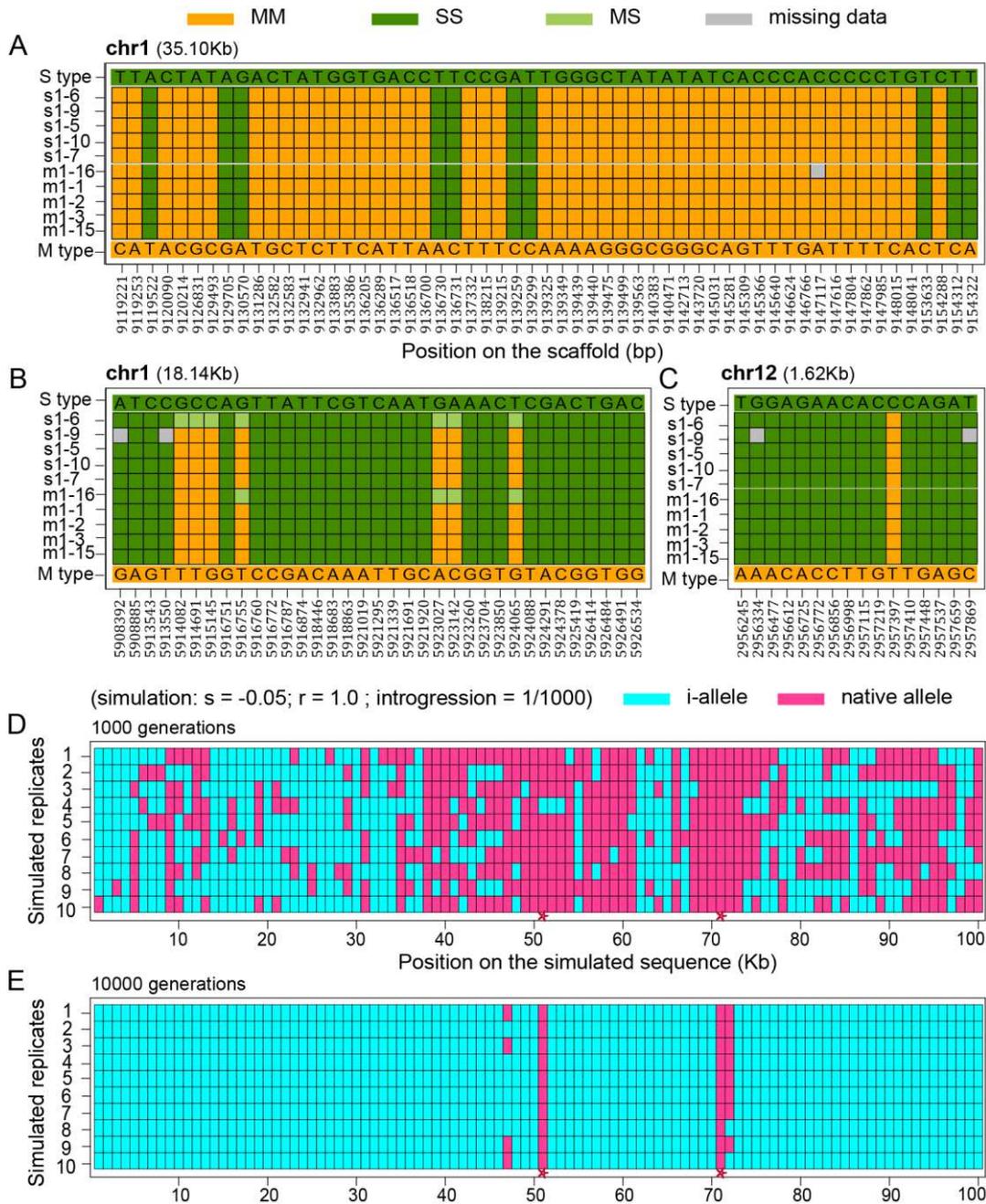
**Figure 3. Key diagnostic characters between *R. mucronata* and *R. stylosa*.** (A) *R. mucronata* (Pulau Burung, Malaysia, 101°50'14.6"E, 2°29'33.7"N) and *R. stylosa* (Zhanjiang, China, 109°45'46.50"E, 21°34'7.32"N) styles from allopatric sites. (B) *R. mucronata* and *R. stylosa* styles from the Daintree River, Australia, sympatric site. (C) Variation of style lengths of *R. stylosa* (green oval) and *R. mucronata* (orange oval) throughout the Indo-West Pacific region. Solid ellipses contrast sympatric samples. (D) Diagrams showing the habitat preferences of *R. mucronata* and *R. stylosa* in a typical estuary (adapted from mangrove ID <sup>43</sup>).



**Figure 4. Admixture of genomes between *R. mucronata* and *R. stylosa*.** (A) Genetic clustering of all 52 individuals of the two species by ADMIXTURE ( $K = 2$ ). Orange and green colors denote the *R. mucronata* (m) and *R. stylosa* (s) components, respectively. Individuals from each population are grouped between adjacent black lines. (B) Distribution of i-allele occurrences in m1 (orange) and s1 (green) populations. Given the five individuals (or 10 haploid genomes) from each population, the occurrence ranges from 0 to 10. Exact numbers are given in the inset table (see also Supplement Fig. S10).

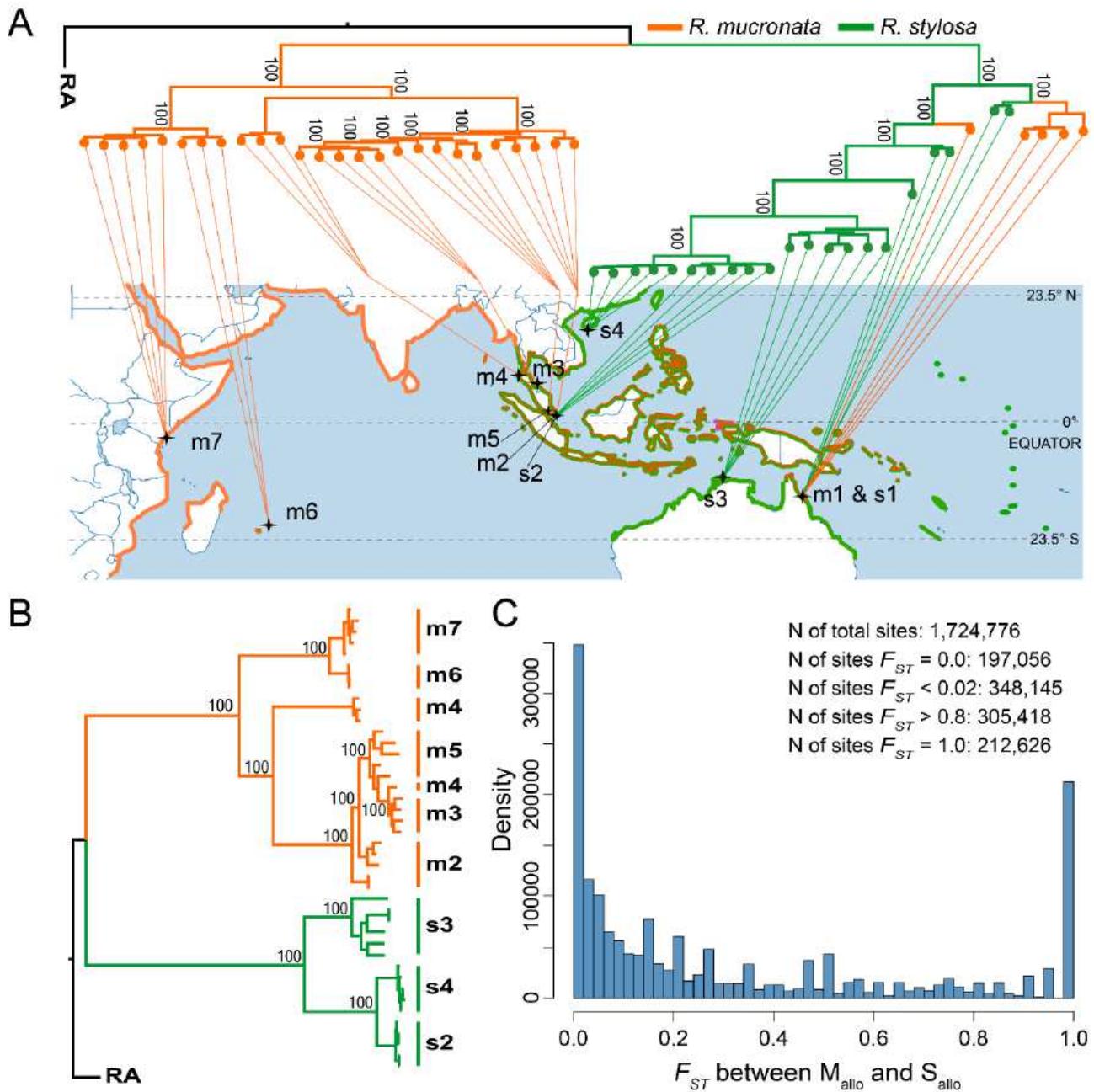


**Figure 5. Interspersion between introgressions and non-introgressions.** (A) A schematic diagram for delineating i-blocks, each of which harbors consecutive i-sites without being interrupted by d-sites. The length of an i-block is determined by the midpoints of the flanking (d, i) intervals. Three i-blocks are shown. (B) The genome-wide landscape of i-blocks. Top 18 longest scaffolds (chr1-18) and the rest of the genome (others) are marked and sibling scaffolds are distinguished by gray rectangles. Each row indicates an individual with each vertical line indicating a site. All 10 individuals from the sympatric s1 and m1 populations are shown. For comparison, one individual is randomly selected from each of other populations (see Fig. S12 for the full display). Each site is color-coded for its genotype: MM (orange), MS (light green) and SS (green) type, where M stands for *R. mucronata* and S for *R. stylosa*. The percentages are summarized on the right. Note that extensive interspersions are observed only in the sympatric samples. (C) A close-up view of i-blocks in the longest scaffold (chr1). Only 10 individuals from the sympatric s1 and m1 populations are shown.



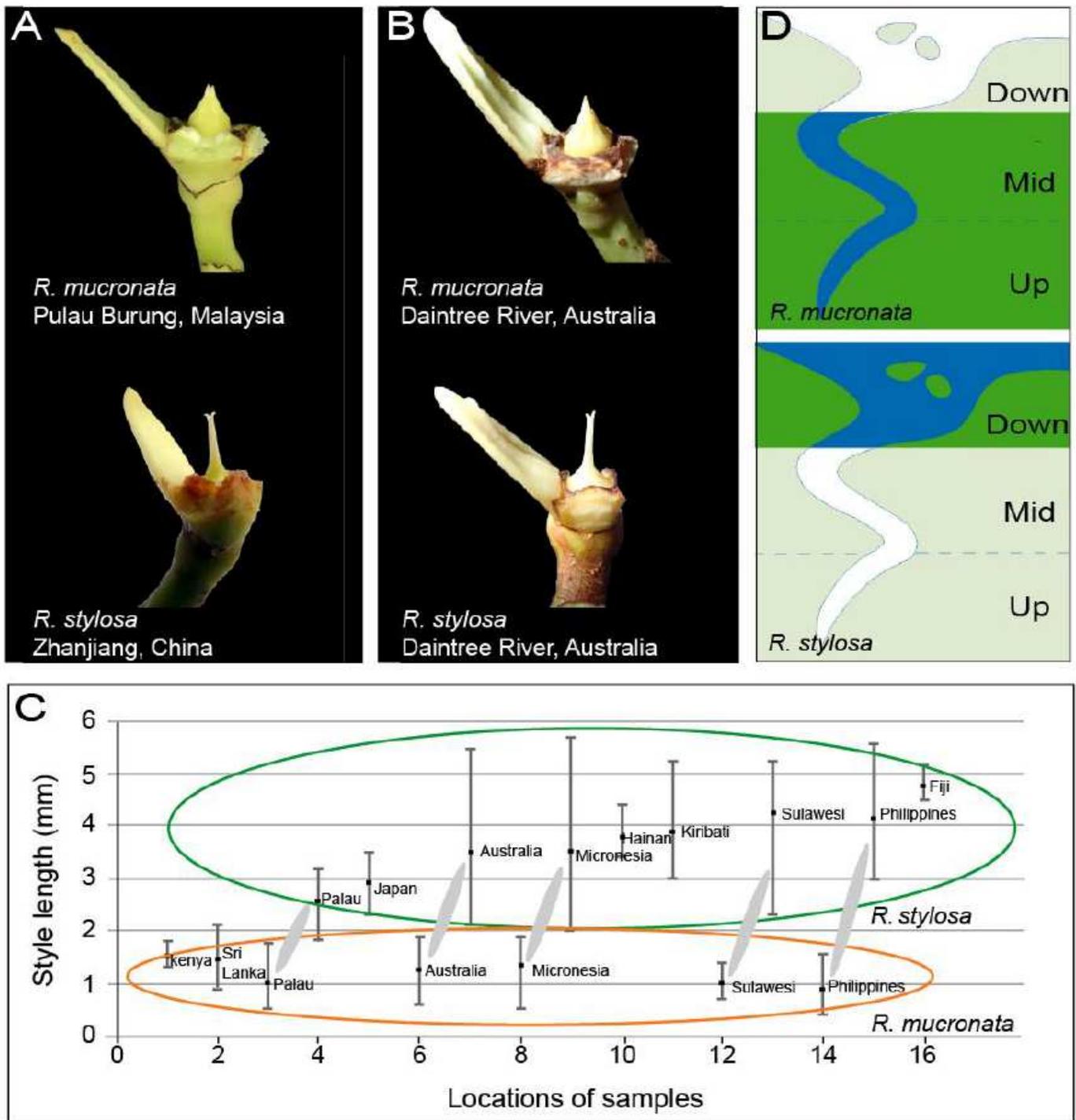
**Figure 6. Examples of i-blocks and simulated introgressions in haploid 100Kb genomes.** (A-C) Examples of i-blocks in m1 and s1 samples at the site level that show the fine-scale delineation of i-blocks. Color codes are the same as in Fig. 4. (D-E) Simulated introgressions in haploid 100 Kb genomes. This example is from a simulation with strong selection ( $s = -0.05$ ), high recombination ( $r = 1.0$  for per 100Kb per generation) and low introgression (1/1000 per generation). Two time points are given (see Materials and Methods and Supplementary Fig. S13 for details). Two speciation genes (or loci) under selection at 51 and 71 Kb are marked by red stars at the bottom. Introgressed and non-introgressed sites are marked in blue and pink, respectively. Note that very fine delineations of blocks are possible under the simulated conditions.





**Figure 2**

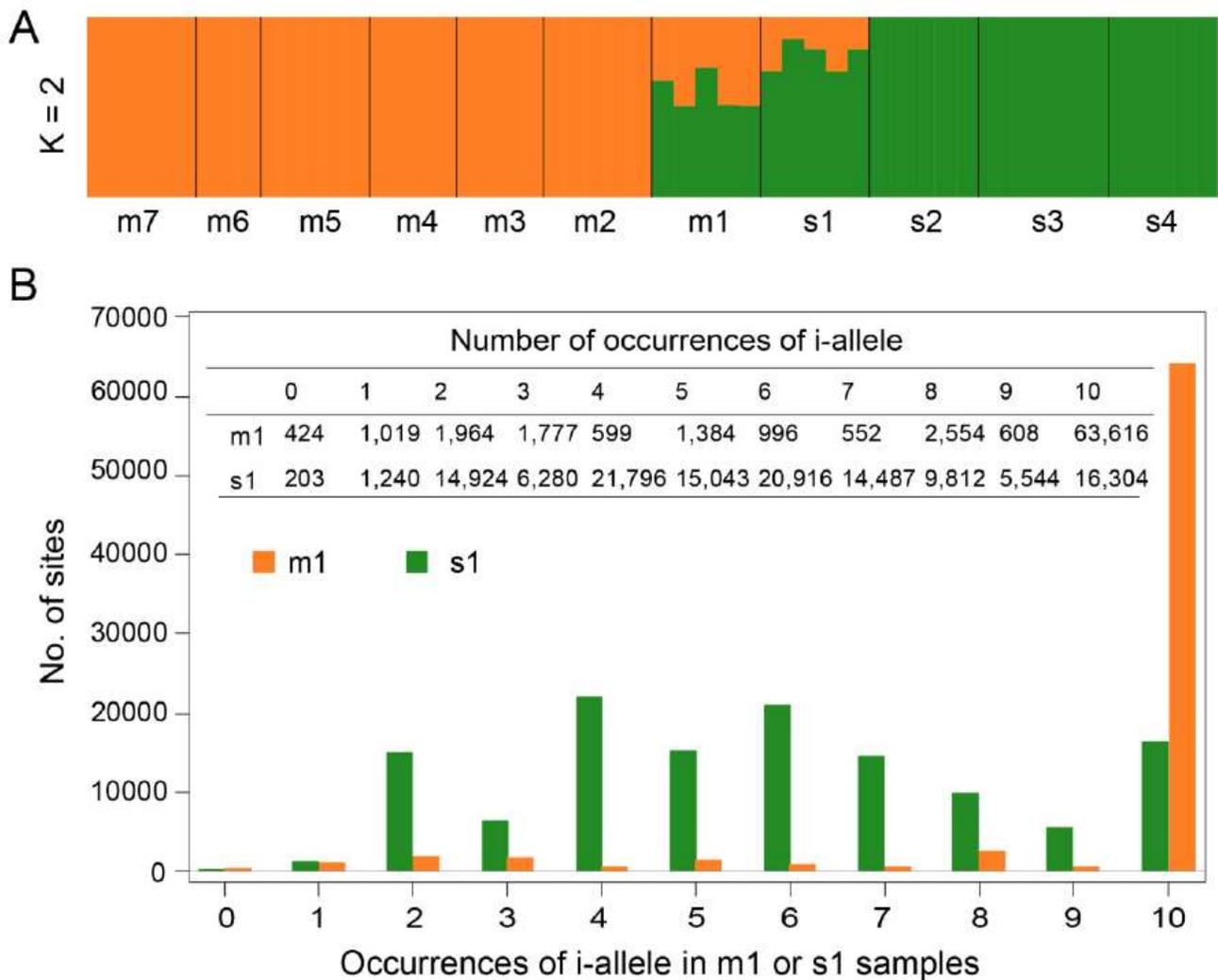
Phylogenetic relationships and biogeographical observations. (A) The biogeography and genealogy of *R. mucronata* and *R. stylosa*. The maximum-likelihood (ML) tree, generated by RAxML with 100 bootstraps, is superimposed on the biogeography. The numbers on the nodes indicate the supporting values. *R. mucronata* is colored in orange while *R. stylosa* is in green. (B) The same phylogeny excluding the sympatric m1 and s1 samples from the Daintree area. We denote the allopatric populations as Mallo (m2-m7) and Sallo (s2-s4). (C) The spectrum of the  $F_{ST}$  statistic between the Mallo and Sallo samples.



**Figure 3**

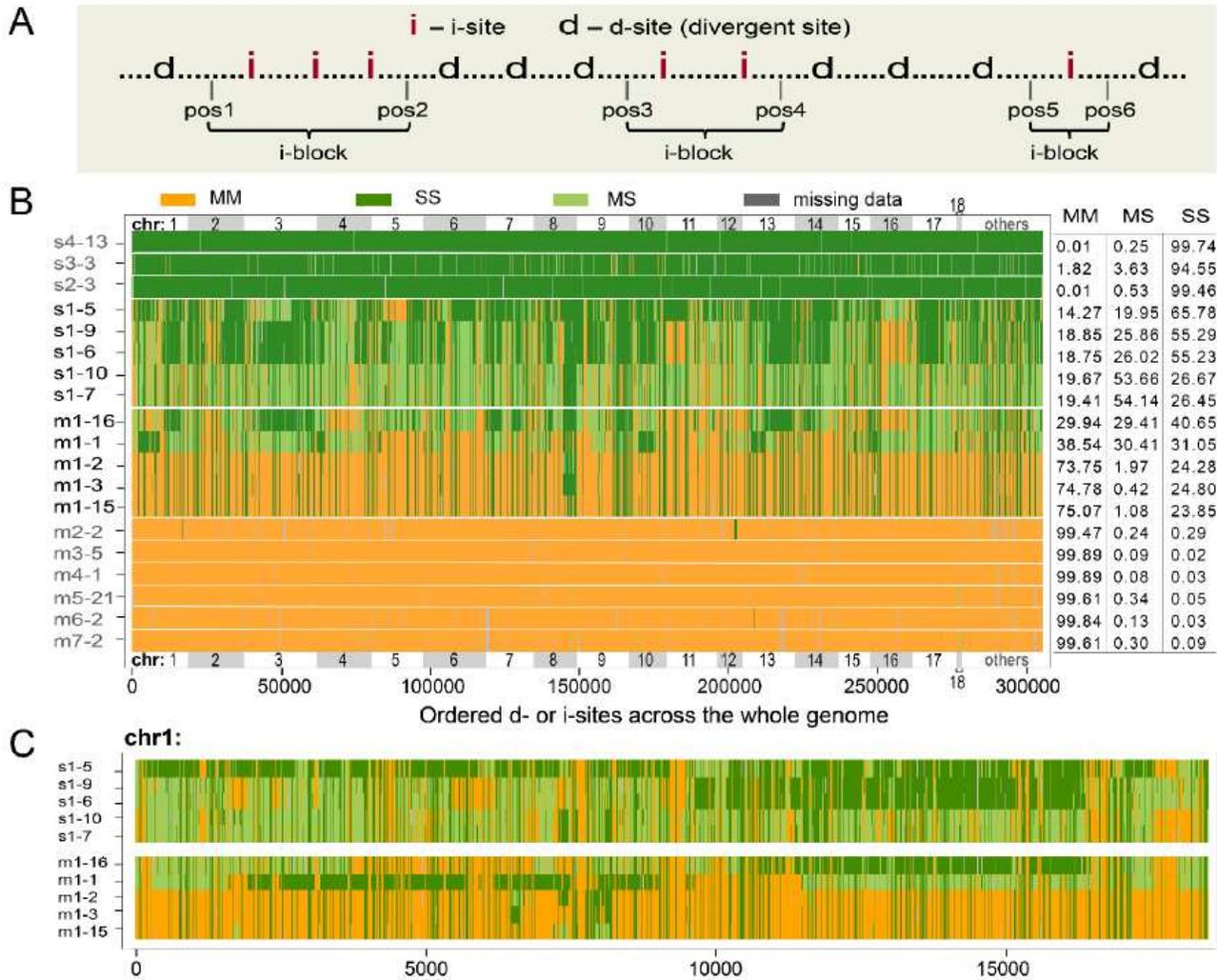
Key diagnostic characters between *R. mucronata* and *R. stylosa*. (A) *R. mucronata* (Pulau Burung, Malaysia, 101°50'14.6"E, 2°29'33.7"N) and *R. stylosa* (Zhanjiang, China, 109°45'46.50"E, 21°34'7.32"N) styles from allopatric sites. (B) *R. mucronata* and *R. stylosa* styles from the Daintree River, Australia, sympatric site. (C) Variation of style lengths of *R. stylosa* (green oval) and *R. mucronata* (orange oval) throughout the Indo-West Pacific region. Solid ellipses contrast sympatric samples. (D) Diagrams

showing the habitat preferences of *R. mucronata* and *R. stylosa* in a typical estuary (adapted from mangrove ID 43).



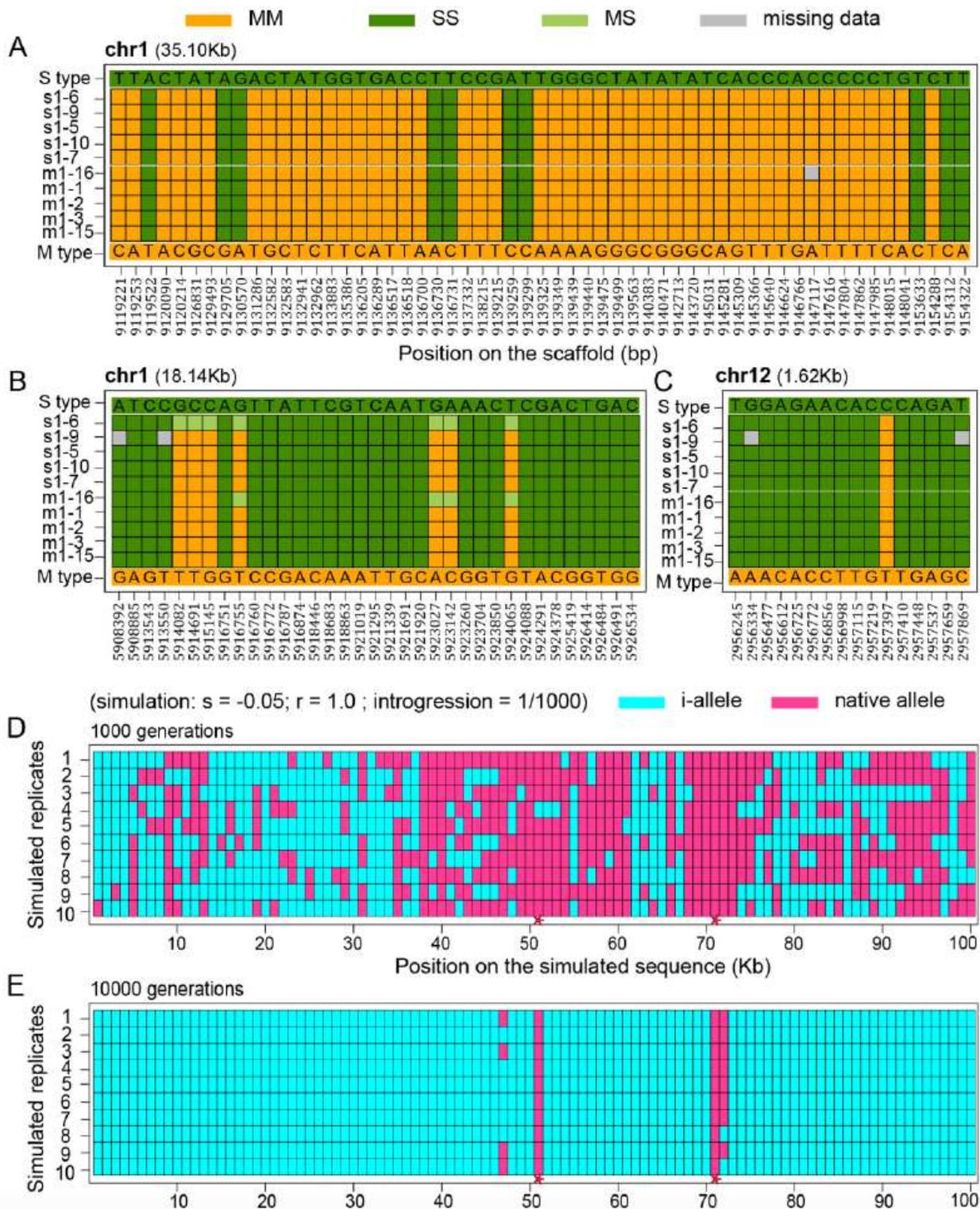
**Figure 4**

Admixture of genomes between *R. mucronata* and *R. stylosa*. (A) Genetic clustering of all 52 individuals of the two species by ADMIXTURE ( $K = 2$ ). Orange and green colors denote the *R. mucronata* (m) and *R. stylosa* (s) components, respectively. Individuals from each population are grouped between adjacent black lines. (B) Distribution of i-allele occurrences in m1 (orange) and s1 (green) populations. Given the five individuals (or 10 haploid genomes) from each population, the occurrence ranges from 0 to 10. Exact numbers are given in the inset table (see also Supplement Fig. S10).



**Figure 5**

Interspersion between introgressions and non-introgressions. (A) A schematic diagram for delineating i-blocks, each of which harbors consecutive i-sites without being interrupted by d-sites. The length of an i-block is determined by the midpoints of the flanking (d, i) intervals. Three i-blocks are shown. (B) The genome-wide landscape of i-blocks. Top 18 longest scaffolds (chr1-18) and the rest of the genome (others) are marked and sibling scaffolds are distinguished by gray rectangles. Each row indicates an individual with each vertical line indicating a site. All 10 individuals from the sympatric s1 and m1 populations are shown. For comparison, one individual is randomly selected from each of other populations (see Fig. S12 for the full display). Each site is color-coded for its genotype: MM (orange), MS (light green) and SS (green) type, where M stands for *R. mucronata* and S for *R. stylosa*. The percentages are summarized on the right. Note that extensive interspersions are observed only in the sympatric samples. (C) A close-up view of i-blocks in the longest scaffold (chr1). Only 10 individuals from the sympatric s1 and m1 populations are shown.



**Figure 6**

Examples of i-blocks and simulated introgressions in haploid 100Kb genomes. (A-C) Examples of i-blocks in m1 and s1 samples at the site level that show the fine-scale delineation of i-blocks. Color codes are the same as in Fig. 4. (D-E) Simulated introgressions in haploid 100 Kb genomes. This example is from a simulation with strong selection ( $s = -0.05$ ), high recombination ( $r = 1.0$  for per 100Kb per generation) and low introgression (1/1000 per generation). Two time points are given (see Materials and Methods and

Supplementary Fig. S13 for details). Two speciation genes (or loci) under selection at 51 and 71 Kb are marked by red stars at the bottom. Introgressed and non-introgressed sites are marked in blue and pink, respectively. Note that very fine delineations of blocks are possible under the simulated conditions.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementStylosapaper9320final.pdf](#)