Speciation with gene flow via cycles of isolation and migration: insights from multiple mangrove taxa

Ziwen He¹, Xinnian Li¹, Ming Yang¹, Xinfeng Wang¹, Cairong Zhong², Norman C. Duke³, Chung-I Wu¹,⁴,⁵, ⁻ and Suhua Shi¹, ⋆

1State 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, Guangzhou 510275, China; 2Hainan Dongzhai Harbor National Nature Reserve Administration, Haikou 571129, China; 3Centre for Tropical Water and Aquatic Ecosystem Research, James Cook University, Townsville 4811, Australia; 4CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China and 5Department of Ecology and Evolution, University of Chicago, Chicago IL 60637, USA

⁎Corresponding authors. E-mails: ciwu@uchicago.edu; lsssh@mail.sysu.edu.cn

Received 19 February 2018; Revised 15 July 2018; Accepted 22 July 2018

ABSTRACT

Allopatric speciation requiring an unbroken period of geographical isolation has been the standard model of neo-Darwinism. While doubts have been repeatedly raised, strict allopatry without any gene flow remains a plausible mechanism in most cases. To rigorously reject strict allopatry, genomic sequences superimposed on the geological records of a well-delineated geographical barrier are necessary. The Strait of Malacca, narrowly connecting the Pacific and Indian Ocean coasts, serves at different times either as a geographical barrier or a conduit of gene flow for coastal/marine species. We surveyed 1700 plants from 29 populations of 5 common mangrove species by large-scale DNA sequencing and added several whole-genome assemblies. Speciation between the two oceans is driven by cycles of isolation and gene flow due to the fluctuations in sea level leading to the opening/closing of the Strait to ocean currents. Because the time required for speciation in mangroves is longer than the isolation phases, speciation in these mangroves has proceeded through many cycles of mixing-isolation-mixing, or MIM, cycles. The MIM mechanism, by relaxing the condition of no gene flow, can promote speciation in more many geographical features than strict allopatry can. Finally, the MIM mechanism of speciation is also efficient, potentially yielding \( m^n \) \((m > 1)\) species after \( n \) cycles.

Keywords: speciation, gene flow, allopatry, geographical isolation, mangrove

SIGNIFICANCE STATEMENT

Mechanisms of species formation have always been a conundrum. Speciation between populations that are fully geographically isolated, or allopatric speciation, has been the standard solution in the last 50 years. Complete geographical isolation with no possibility of gene flow, however, is often untenable and is inefficient in generating the enormous biodiversity. By studying mangroves on the Indo-Malayan coasts, a global hotspot of coastal biodiversity, we were able to combine genomic data with geological records on the Indo-Pacific Barrier that separates Pacific and Indian Ocean coasts. We discovered a novel mechanism of speciation that we call mixing-isolation-mixing (MIM) cycles. By permitting intermittent gene flow during speciation, MIM cycles can potentially generate species at an exponential rate, thus combining speciation and biodiversity in a unified framework.

INTRODUCTION

Speciation driven by geographical isolation with no possibility of gene flow, or strict allopatric speciation, has been the standard model of neo-Darwinism [1,2]. Although occasional exceptions are acceptable to this view [3–5], extensive violations of strict allopatry would contradict many of its central tenets. One of these tenets is the nature of species as defined by the Biological Species Concept [2]. The argument for strict allopatry has usually been that gene flow would homogenize diverging populations and retard speciation [2]. After the completion of speciation, secondary contact may lead to a hybrid zone but the process of speciation should have become irreversible by then [6–8].

The stringent requirement for complete geographical isolation, however, is not without difficulties. Chief among them is the paucity of geographical features that can fully stop gene flow to sustain
long-term isolation. As a result, the observed biodiversity seems too extensive to rely solely on the limited opportunities for strict allopatric speciation [9]. Fisher outlined a verbal theoretical model of clinal speciation [10] and Endler suggested that parapatric speciation, arising between adjacent populations that continue to exchange genes at a reduced level, may be far more common than allopatry [11]. Divergent selection in parapatry can be sufficient to overcome the homogenizing effects of migration if individual genic effects are examined (see [6]). In this genic view, the level of divergence at the completion of speciation would be uneven across the genome. In particular, there may exist ‘genomic islands of speciation’ (GIS) that are involved in divergent adaptation or reproductive isolation [5,12–14].

Evidence for locus-dependent gene flow leading to the formation of GIS has been widely reported [5,12–14], although Cruickshank and Hahn rejected many reported examples of GIS as products of processes unrelated to speciation [15]. More generally, it has been pointed out that genomic data alone could not have the power to reject the allopatric model, even when GIS can be properly identified [16]. In particular, if geographical isolation arises between subdivided populations, allopatry would likely be falsely rejected. Other sources of data are therefore necessary.

The resolution of the issue of speciation with gene flow may be possible if historical data on geographical barriers, offering a temporal perspective, are available. Fauna and flora of the two ocean coasts delineated by the Indo-Pacific Barrier (IPB) are particularly suited to such inquiries (Fig. 1a). The Strait of Malacca, a main feature of the IPB, can impose large-scale geographical isolation for taxa with ocean-current-dependent migration. Unlike the isolation at the Isthmus of Panama, the IPB isolation is not permanent. When the sea level rose and fell periodically during the Pleistocene [17], the Strait of Malacca, which is much shallower than the two oceans, closed and opened intermittently to ocean currents and gene flow (Fig. 1b) [18]. The timing of the alternation of the phases can be inferred from geological records (Fig. 1b). Hence, the DNA divergence pattern can be superimposed on the geographical records of the physical barrier itself.

A larger issue raised by speciation mechanisms concerns biodiversity. There are a number of biodiversity centers globally. Among them, an exceptionally biota-rich region is found along the Indo-Malayan coasts [1,19]. Major groups of flora and fauna display unequalled species richness on these coasts [20–22]. Mechanisms of speciation have been proposed, and rejected, as an explanation for exceptionally high diversity in such hot-spots [23], for two reasons. First, these centers often do not have geographical features that can facilitate allopatric speciation by imposing long-term geographical isolation [19]. Second, speciation by strict allopatry (e.g. at the Isthmus of Panama) is not an efficient mechanism to generate the multitude of species because species would simply double in number. This study will attempt to connect speciation mechanisms and species richness in a single framework. Finally, given the breadth of the subject matter, necessary backgrounds and potential criticisms cannot be fully addressed in the main text. These additional topics are therefore presented in the section ‘ Replies to objections to the MIM model’ of the Supplementary Note, available as Supplementary Data at NSR online.

RESULTS

In this study, we analyse the divergence in five distantly related taxa of mangroves, which are woody plants that independently invaded the intersection between land and sea within the last 100 million years (Myrs) [21,24–26]. Because mangroves occupy a narrow band along tropical coasts, their distributions are essentially one-dimensional, making it easier to identify geographical barriers between species. For mangroves along the two ocean coasts (referred to as the East vs. West regions in Fig. 1a), the barrier is often the Strait of Malacca, which opens and closes periodically to ocean currents, and serves as a conduit for mangrove seed dispersal when open (see ‘Introduction’). Globally, there are 70 or so mangrove species and >80% of them can be found along the Indo-Malayan coasts [27]. Many of these mangrove taxa have existed and undergone diversification only in this region. In contrast, only eight species exist in the New World tropics [21,24,28]. Since other taxa are also highly diverse on the Indo-Malayan coasts [29,30], the geographical mechanism of speciation in mangroves may be broadly applicable to other fauna and flora in this region.

In this study, we approach mangrove speciation from both ends: divergence between good species and differentiation between geographical populations. By doing so, we resolve the dilemma in studying speciation. The dilemma is that good species may be too divergent to inform about speciation events [12,15,17,31], but subspecies and geographical populations are not, and may not become, true species.

We have generated high-quality whole-genome sequences from multiple species of mangroves [26]. For the analysis of speciation history, many more samples but fewer genes per species are necessary to study population differentiation. The
Figure 1. Geography and patterns of speciation in the Indo-Western Pacific. (a) Population samples were collected from both the Indian and Pacific Ocean coasts, separated by the Strait of Malacca. Red dots mark sampling sites; H, Hainan Island; G, Gulf of Thailand; W, West coast of Thailand. E stands for H + G to the East of the Strait. The red arrow-headed line depicts ocean current (and potential gene flow) through the Strait of Malacca between the Indian and Pacific coasts. (b) Sea-level changes in the last 2.5 Myrs. The red line marks the depth of the Strait of Malacca (~25 meters). (c) Ceriops; (d) Sonneratia; (e) Rhizophora. The species distribution of each genus is based on Mangrove ID [33]. Species pairs delineated by the Strait of Malacca are shown in red and green, while the phylogeny is given in the inset. Additional species boundaries are given in Supplementary Fig. 1, available as Supplementary Data at NSR online. (f) Genetic divergence in 50-Kb sliding windows across the genome between R. mucronata and R. stylosa. Alternating colors denote different scaffolds; the dotted horizontal line marks the highest 5th-percentile in divergence. Red points indicate peaks of consecutive windows with elevated divergence (~100 Kb). The four red arrows indicate divergence peaks that remain after controlling for mutation rate variation, as scaled by the divergence between R. apiculata and R. mucronata.
geographical populations of five common mangrove species on the two coasts to the East and West of the Strait are shown in Fig. 1a. Sampling was done in 14 locations of the three areas: Hainan Island in China (H), the East coast of Thailand facing the gulf (G) and the west coast of Thailand (W). In total, approximately 1700 individuals from the five species were collected (Supplementary Table 1, available as Supplementary Data at NSR online) and subjected to sequencing. Following a published method [32], we obtained an average of 70 Kbps of sequence across 80 genes per individual (Supplementary Table 2, available as Supplementary Data at NSR online).

**Speciation history in the Indo-Western Pacific (IWP)**

The Indo-Malayan coasts, as part of the IWP and depicted in Fig. 1a, represent an important biodiversity hotspot. Near the tip of the Malay Peninsula, more than 20 mangrove species can be found in a local population ([33,34] and our field observations). At least nine mangrove genera had formed relatively recently originated species along these coasts and the recent speciation events (< 4% divergence) are shown in Table 1. The five genera most actively involved in recent speciation are analysed in detail here. Documented hybridization is not uncommon in areas of sympatry (Table 1). Molecular typing has shown that all hybrids belong to first-filial generation [35–40] and exhibit poor pollen maturation or seed germination in the hybrids in planting experiments [41,42].

The Strait of Malacca is a major geographical barrier for mangroves in the IWP (Fig. 1b and c). One example concerns the species pair Ceriops decandra and C. zippeliana (Fig. 1c), between which the ex tant boundary is right along the Strait. The species boundary between Sonneratia owaya and S. griffithii is broader but also falls along the Strait (Fig. 1d), whereas the ranges of Rhizophora mucronata and R. stylosa (Fig. 1e) overlap broadly along the two ocean coasts adjacent to the Strait of Malacca, although their general distributions suggest post-speciation dispersion across the Strait, which is in keeping with the better dispersal ability of Rhizophora compared to either Ceriops or Sonneratia [43]. Three other genera are likely to have experienced post-speciation migration through the Strait of Malacca, much like Rhizophora. They concern the species pairs: Avicennia rumphiana and Av. alba; Lumnitzera littorea and L. racemosa; and Bruguiera sexangula and B. gymnorhiza (Supplementary Fig. 1, available as Supplementary Data at NSR online). The geology of the region and the sea-level records are shown in Fig. 1a and b, and indicate that the East and West regions would be strongly isolated when the sea level drops below ~25 meters, which is the historical norm.

It is important to point out that the Strait of Malacca connecting/separating the Pacific and Indian Ocean coasts is only one of many barriers in the IWP. Other geographical barriers can also be inferred. For example, the Torres Strait may have restricted the distributions of the sibling species Sonneratia caseolaris and S. lanceolata in northern Australia (Supplementary Fig. 1a, available as

<table>
<thead>
<tr>
<th>Genus</th>
<th>No. of species</th>
<th>No. of hybrids</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Estimated divergence time (Myrs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ceriops</td>
<td>5</td>
<td>1</td>
<td>0.28</td>
<td>2.10</td>
<td>0.97 ~ 7.23</td>
</tr>
<tr>
<td>Rhizophora</td>
<td>3</td>
<td>3</td>
<td>0.28</td>
<td>1.50</td>
<td>1.20 ~ 6.44</td>
</tr>
<tr>
<td>Sonneratia</td>
<td>6</td>
<td>4</td>
<td>0.70</td>
<td>3.75</td>
<td>1.25 ~ 6.67</td>
</tr>
<tr>
<td>Avicennia</td>
<td>5 (7)</td>
<td>1</td>
<td>0.80</td>
<td>3.70</td>
<td>1.31 ~ 6.07</td>
</tr>
<tr>
<td>Aegiceras</td>
<td>2</td>
<td>0</td>
<td>2.14</td>
<td></td>
<td>1.66</td>
</tr>
<tr>
<td>Xylocarpus</td>
<td>3</td>
<td>0</td>
<td>1.88</td>
<td></td>
<td>2.00</td>
</tr>
<tr>
<td>Lumnitzera</td>
<td>2</td>
<td>1</td>
<td>2.72</td>
<td></td>
<td>2.85</td>
</tr>
<tr>
<td>Kandelia</td>
<td>2</td>
<td>0</td>
<td>0.78</td>
<td></td>
<td>2.23</td>
</tr>
<tr>
<td>Bruguiera</td>
<td>6</td>
<td>1</td>
<td>0.99</td>
<td></td>
<td>4.63</td>
</tr>
</tbody>
</table>

The divergence level was estimated from 3–60 loci in each genus. Sequences were obtained from published studies or by ourselves and deposited in GenBank (see Supplementary Table S, available as Supplementary Data at NSR online).

In millions of years (Myrs). Divergence times were calculated as genetic divergence divided by the corresponding nucleotide substitution rate. The rate for Lumnitzera and Xylocarpus was estimated as 4.78 x 10^{-9} /site/year based on internal transcribed spacers of the nuclear ribosomal DNA [77]. We obtained whole-genome and/or whole-transcriptome data for the other genera from a separate study ([26] and He et al., unpublished data). The substitution rate for each genus was then inferred for the specific branches using PhyML [91] and PAML [92] in conjunction with fossil dating. A further adjustment was made to compensate for the different substitution rates between coding and non-coding regions (see Supplementary Table 6 and Supplementary Note, available as Supplementary Data at NSR online).

If the three subspecies of Av. marina are counted separately, the number would be 7.

Divergence information is available for only one pair of species in this genus.
Supplementary Data at NSR online; reviewed in [44]). The biodiversity in the IWP in relation to these barriers will be discussed below.

**Speciation with gene flow between the two ocean coasts**

The time of species divergence in the nine genera listed in Table 1 was estimated for each node of the phylogenetic tree based on DNA-sequence data and the estimated species-specific nucleotide substitution rates (see Supplementary Note, available as Supplementary Data at NSR online). In eight of these nine genera, the most recent species divergence time is within the last 3 Myrs. The oldest divergence time in Table 1 is about 6.5 Myrs ago. The most recent events within each genus generally fall in the time frame depicted in Fig. 1b, which shows the possible periods of gene flow (above –25 meters, indicated by the red broken line).

A history of gene flow should be reflected in the genomic data because genomic segments involved in differential adaptation (in physiology, morphology, reproduction, etc.) should be more divergent than the rest of the genome [6,45]. Many statistical tests have been developed to test this hypothesis by asking whether the level of divergence is ‘over-dispersed’ across the genome. Here, we employed two methods (Supplementary Fig. 2, available as Supplementary Data at NSR online) on *R. mucronata* and *R. stylosa* (Fig. 1e), using the complete genome sequences published recently [26]. In the first method, the divergence level in genomic regions is compared with that of intergenic regions. The second method [47] relies on the variance in divergence across the genome. Both methods implement likelihood-ratio tests to compare the allopatric (*H₀*) and speciation with gene flow (*H₁*) models. In both methods, the null model is rejected with high confidence (*P ~ 0*; Supplementary Table 3, available as Supplementary Data at NSR online), thus supporting the model of gene flow during speciation (see details for Supplementary Note, available as Supplementary Data at NSR online). In order to identify the genomic segments most likely involved in speciation, we conducted a sliding-window analysis. Very large GIS regions between *R. mucronata* and *R. stylosa* that are unusually divergent are shown in Fig. 1f (see legend). Four of them, marked by red arrows, are more stringently called. In total, 40 GIS segments are identified for a total of 4775 Kb, or 2.33% of the sequenced genome.

Figure 1f follows the standard procedure in testing ‘speciation with gene flow’ and rejects the null hypothesis. However, Yang *et al*. recently suggested that the statistical rejection is valid only for the simplest form of allopatry. For example, if gene flow occurs between geographical populations before, but not during, speciation, the null model would still be rejected, hence leading to the false rejection of allopatry [16,48]. In other words, the tests are done because the failure to reject would be biologically informative while the rejection is much less so. In cases of rejection, other types of data (e.g. geographical distributions of species and the nature of the putative barrier prior to the completion of speciation) need to be superimposed on the genomic analyses. In the remaining sections, such data will be used on geographical populations located along the two ocean coasts. The objective is to estimate the minimal time required for speciation, which will then be compared with the geological records of the geographical barrier itself.

**Differentiation between geographical populations on the two ocean coasts**

The Strait of Malacca has served as a geographical barrier leading to speciation in the past. We asked whether it continues to serve as a barrier for geographical differentiation at present. Morphological observations support the inference of East–West differentiation (see Fig. 2a and b) and DNA-sequence divergence provides the time depth of geographical differentiation. The latter is usually expressed by partitioning the diversity within and between regions. Both π₉, the genetic diversity within each area (H₉, G or W), and π₆, the total diversity of the species, are listed in Table 2 and legends, as well as Supplementary Table 4 and Supplementary Fig. 3, available as Supplementary Data at NSR online. Population divergence between regions, denoted by $F_{ST} = (\pi_{T} - \pi_{R})/\pi_{T}$ [49], generally follows the speciation pattern.

One of the five species, *Ceriops tagal* has unusually low diversity ($\pi_{T} = 0.343/Kb$, less than $\frac{1}{2}$ of that of the next lowest species) and hence little differentiation among all populations. Table 2 shows that all other species exhibit a larger $\pi_{T}$ than $\pi_{R}$ and strong population differentiation. Figure 2c shows pairwise differentiation patterns between the three geographical areas. The divergence is relatively low in the H–G comparison in the three species with intermediate diversity (*Rhizophora apiculata*, *Sonneratia alba* and *Avicennia marina*), despite substantial geographical
Figure 2. Geographical differentiation in morphology and DNA sequence. (a) and (b) Morphological differentiation of *Ae. corniculatum* from the East (a: samples from Hainan, China) and West (b: samples from Ranong, Thailand) regions. (c) Box plots of the $F_{ST}$ statistic for each of the five species. For the three species with intermediate genetic diversity (*R. apiculata*, *S. alba* and *Av. marina*), $F_{ST}$ between H and G is lower than the other two. (d) to (f) Examples of haplotype networks in the same three species that show strong East–West divergence.

Table 2. Differentiation rate and time between East and West populations.

<table>
<thead>
<tr>
<th>Samples and sequences</th>
<th>C. tagal</th>
<th>R. apiculata</th>
<th>S. alba</th>
<th>Av. marina</th>
<th><em>Ae. corniculatum</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of stands (H, G, W)$^a$</td>
<td>2, 1, 1</td>
<td>3, 2, 1</td>
<td>2, 2, 3</td>
<td>2, 3, 1</td>
<td>3, 1, 2</td>
</tr>
<tr>
<td>Sample size (H, G, W)</td>
<td>200, 100, 100</td>
<td>90, 68, 34</td>
<td>185, 100, 150</td>
<td>200, 89, 35</td>
<td>174, 50, 87</td>
</tr>
<tr>
<td>No. segments (Total Kb)</td>
<td>102 (76.6)</td>
<td>124 (65.2)</td>
<td>101 (59.6)</td>
<td>150 (85.2)</td>
<td>115 (57.3)</td>
</tr>
<tr>
<td>Differentiation between East and West regions</td>
<td>$F_{ST}$ ($\pi_T, \pi_R$)$^b$ (/Kb)</td>
<td>0.267</td>
<td>0.48</td>
<td>0.67</td>
<td>0.72</td>
</tr>
<tr>
<td></td>
<td>(0.343, 0.251)</td>
<td>(1.241, 0.645)</td>
<td>(2.151, 0.701)</td>
<td>(2.519, 0.705)</td>
<td>(10.746, 2.640)</td>
</tr>
<tr>
<td>Estimation of mutation rate and divergence</td>
<td>Mutation rate, $\mu$ (/Kb/generation)$^c$</td>
<td>$1.80 \times 10^{-5}$</td>
<td>$1.63 \times 10^{-5}$</td>
<td>$2.84 \times 10^{-5}$</td>
<td>$3.26 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>$N_e \mu$ in SIM (MIM)$^d$</td>
<td>0.009 (0.007)</td>
<td>0.029 (0.035)</td>
<td>0.035 (0.030)</td>
<td>0.020 (0.019)</td>
</tr>
<tr>
<td></td>
<td>$N_e m$ in SIM (MIM)$^d$</td>
<td>0.348 (0.420)</td>
<td>0.796 (0.282)</td>
<td>0.541 (0.183)</td>
<td>0.265 (0.091)</td>
</tr>
<tr>
<td></td>
<td>$T_{min}$ (Myrs)</td>
<td>0.18</td>
<td>1.38</td>
<td>1.14</td>
<td>1.10</td>
</tr>
<tr>
<td></td>
<td>$P^e$</td>
<td>0.67</td>
<td>&gt;0.999</td>
<td>&gt;0.999</td>
<td>&gt;0.999</td>
</tr>
</tbody>
</table>

$^a$H, Hainan; G, Gulf of Thailand; W, West coast of Thailand (see Fig. 1a).

$^b$\(\pi_T\) is the average within-region diversity, \(\pi_R\) is the total diversity and $F_{ST} = (\pi_T - \pi_R)/\pi_T$ (see text). Nucleotide diversities for population stands ($\pi_S$) are given in Supplementary Table 4, available as Supplementary Data at NSR online.

$^c$Mutation rate estimation is based on fossil record divergence time (see legends in Table 1).

$^d$\(N_e \mu\) and \(N_e m\) are estimated separately for the SIM and MIM models (see text and Supplementary Note, available as Supplementary Data at NSR online).

$^e$Maximum likelihood probability that the MIM cycles model is better than the SIM model, based on 2000 replicate simulations.

Geographical differentiation can be analysed in greater detail by analysing haplotype structures. Three examples of haplotype networks are shown in Fig. 2d–f (see more cases in Supplementary Fig. 5, available as Supplementary Data at NSR online). The haplotypes can be clearly divided into two clades, referred to as the Eastern or
Western haplotype, depending on where they are more commonly found. The existence of distinct haplotypes without intermediates usually indicates strong population differentiation [50]. Both the FST statistics and haplotype structures hence suggest strong differentiation between the East and West regions demarcated by the Strait of Malacca.

**DNA-sequence divergence vs. geological record: how much time is needed for mangrove speciation?**

Under the past sea-level changes [17], the East and West regions have experienced cycles of isolation and admixture due to the repeated opening and closing of the Strait (see Fig. 1b). To compare the geological records of barrier duration with the divergence history inferred from genomic sequences, it is necessary to estimate the time required for speciation to be completed (T_{app}, or speciation time). This can then be compared to the isolation time (T_{iso}), the length of the periods when physical barriers to gene flow were recorded in historical data.

If we assume strict allopatry (Fig. 3a), speciation needs to be completed during geographical isolation, or T_{app} < T_{iso}. From Table 1, species divergence takes 1.2–6.7 Myrs, with a mid-point T_{app} of ~4 Myrs. (The lowest estimate of 0.84 Myrs in Ceriops is less reliable due to its very low mutation rate; see Table 2.) From Fig. 1b, T_{iso} is always smaller than 0.5 Myrs and rarely larger than 0.2 Myrs. Obviously, the allopatric condition of T_{app} < T_{iso} is not met. Nevertheless, since the divergence time between good species given in Table 1 represents overestimation of T_{app}, the rejection of T_{app} < T_{iso} is not informative.

We shall now use the lower bound estimate of T_{app} against T_{iso}. This lower bound is the divergence time between geographical populations. A new population genetic framework is developed for the purpose of estimating T_{app} between two randomly chosen genes from the same or different populations. This new framework is presented in detail in Supplementary Note, available as Supplementary Data at NSR online. It is distinct from previous models because it will be needed to compare the allopatric model (Fig. 3a) with our new MIM model (Fig. 3c) with multiple cycles of isolation and migration. The likelihood of observing various distributions of divergence is formulated as the function of T_{app}, N_{e} and m; where N_{e} is the effective population size and m is the migration rate (Table 2). We then use the maximum-likelihood estimates (MLE) to obtain parameters (Table 2). Note that the null model here is strict allopatry, portrayed by the single isolation-mixing (SIM) cycle (Fig. 3a). If gene flow occurred during isolation, we would underestimate T_{app} and the rejection of allopatry would be conservative.

Figure 3b presents the estimated T_{app} for the five species of mangroves under the allopatric SIM model. For a comparison, the temporal sequence of migration and isolation phases at the Strait of Malacca is also shown. With the exception of C. tagal, the estimated T_{app} values exceed 1.2 Myrs in the four other species. As the null model of T_{app} < T_{iso} is rejected, T_{app} must span several cycles of isolation-mixing (see Fig. 3b).

**Speciation through MIM cycles**

Speciation in mangroves along the Pacific vs. Indian Ocean coasts had to go through cycles of isolation interspersed by episodes of gene flow, as recorded in the geological data (Fig. 3b). This mode of speciation will be referred to as the MIM model. The likelihood-ratio test (last row of Table 2) shows that the MIM model agrees with the observations better than the SIM model (Supplementary Note, available as Supplementary Data at NSR online), except in C. tagal, which has a very low species-wide polymorphism.

Under the MIM model, the distribution of neutral divergence among loci should be broader than under SIM, if everything else is equal. We use D_{max} (differences between the two most divergent haplotype at any locus) as the measure. The distribution of D_{max} is given in Fig. 4a–c. The vertical red lines represent the average level of divergence between species or subspecies. All three species have
Figure 4. Patterns of genetic divergence under SIM vs. MIM model of speciation. (a) Divergence among *R. apiculata* populations. The distribution of \(D_{\text{max}}\) (differences between the two most divergent haplotypes for each gene) across loci is depicted. The bars add up to the total percentage of loci that show East–West divergence (given next to the species name). The vertical red lines indicate the level of species (solid line) and subspecies (dotted line) divergence. Note that the divergence between the geographical populations often exceeds that of subspecies, or even species. Inset figures present the standard deviations of \(D_{\text{max}}\) simulated under both MIM and SIM models. The observed value, indicated by the red arrow, is in agreement with the MIM (green line) but not with the SIM model (grey line). (b) *S. alba*. (c) *Av. marina*. (d) Changes in demography (population growth and differentiation) represented by the evolution of ‘effective population size’. The PSMC method [51], when applied to the whole-genome sequencing data, can reveal changes in demography through time. Here, two individuals from each species were used, portrayed by a solid and a dotted line, respectively. Because the effective population size is sensitive to population subdivision, the analysis can discriminate between the SIM and MIM models. As shown, the population size increases gradually back in time, which is the characteristic pattern for the MIM model. In contrast, the SIM model would yield an extremely steep increase. (e) Three scenarios of divergence and eventual speciation. Blue shades indicate periods of migration that punctuate long periods of isolation. Speciation is indicated by high divergence. In the conventional view (black line), gene flow would reverse the divergence. Under the MIM model, the level of divergence is only minimally affected by gene flow (red line). Some parameter combinations under the MIM model would underlie a third possibility (dotted blue line) whereby gene flow after a period of isolation may speed up divergence (see text).

Many loci where \(D_{\text{max}}\) is larger than the level of (sub-)species divergence (upper panels). These loci may reflect aspects of the East–West divergence due to geographical isolation. The standard deviations of \(D_{\text{max}}\) are simulated and plotted (insets in Fig. 4a–c). The observations are indeed much larger than the predictions of the SIM model and fall within the simulated distributions under the MIM cycles. Thus, the divergence of mangroves on these coasts may have been influenced by periodic gene flow increasing among-locus variation. Because isolation increases genetic variation, it also increases the effective population size. Hence, MIM and SIM models would show distinct
patterns. As the genomes of three of the five species have been sequenced ([26] and He et al., unpublished data), we re-sequenced two additional individuals for each of the three species. The pairwise sequentially Markovian coalescent (PSMC) method [51] infers effective population sizes at different time points in the past by comparing haploid genomes. Periods of isolation are reflected in non-coalescence and can be defined as changes in effective population size.

The PSMC results on *R. apiculata*, *S. alba* and *Av. marina* are given in Fig. 4d. While PSMC is usually used to model the changes in population sizes, we use it here in the context of the timing of population differentiation on the Pacific vs. Indian Ocean coasts (see Supplementary Note, available as Supplementary Data at NSR online). All three species show very small effective population sizes in the last 20,000 years, corresponding to the retreat of the last global glaciation. Going back in time, the effective population sizes increase gradually, suggesting isolated populations that have had low or intermittent gene flow during the preceding 2 Myrs. The overall PSMC patterns indicate historical gene flow spread over a long span of time, in accordance with the geological records. Had the gene flow been concentrated in a short period, the simulated SIM model would yield a steep increase in effective population size during a very short window of time (Fig. 4d).

**DISCUSSION**

Gene flow is conventionally perceived as a homogenizing force that can reverse population divergence and block speciation (black line in Fig. 4e). This has been the principal consideration of the strict allopatric model of speciation. The absence of gene flow due to geographical isolation is eventually superseded by the evolution of reproductive isolation that underpins the Biological Species Concept [2,52]. In recent years, the genic perspective suggests that gene flow during speciation would not necessarily impede divergence, as long as selection is not swamped by migration (red line in Fig. 4e) [6,12,53,54]. By superimposing the genomic information on the geological records, this study demonstrates that speciation on the Indo-Malayan coasts must have progressed in alternate phases of gene flow and isolation.

The MIM model therefore bridges two large sets of speciation literature. In one set, the main concern has been about the geological and phylogeographical records of speciation, which have been expertly reviewed by Hewitt [55]. It is, however, not clear whether the phylogeographical literature has rejected the model of strict allopatry or has reinforced it. For example, depending on when a hybrid zone is formed, geographical records may either suggest ‘speciation with gene flow’ or reinforce the view that a hybrid zone would reverse divergence until after speciation is *fait accompli* [56,57]. In this backdrop, earlier cyclic hybridization models linking climatic cycles with speciation [58–60] are extensions of the allopatric model. In these extensions, speciation is completed in one cycle with full isolation followed by migration. The process would continue through cycles of geographical speciation and post-speciation range expansion.

A second set of the literature concerns the genomic divergence that can reveal the speciation history [6,12,14,61,62]. Nevertheless, as shown by several analyses [16,48], genomic data can inform about the occurrence of gene flow but not about when it happened. Gene flow prior to the onset of speciation might be misinterpreted to be gene flow during speciation. No less important, gene flow could be a continuous trickle or might be concentrated in short episodes of geographical panmixia, interspersed with periods of strict isolation. These isolation phases are important for the evolution of postmating reproductive isolation because incompatibility cannot evolve easily under gene flow [63,64]. In this sense, the MIM model has features of both allopatry and ‘speciation with gene flow’.

Interestingly, it has been postulated that gene flow may even speed up speciation (the blue dotted line in Fig. 4e). This could happen if and when adaptive gene complexes, built up during isolation, are shuffled to generate many new combinations. Hybrid speciation [65–67] and adaptive radiation by hybrid swarms are such examples [68]. Furthermore, many domesticated breeds were indeed created by hybridization between existing varieties [69–71]. Thus, both plant and animal domestication resembles the MIM cycles, whereby breeds were separately domesticated with occasional exchange of genes. Although the idea of well-timed gene flow speeding up speciation is attractive, there is currently no evidence that it applies to mangrove speciation.

Finally, the MIM model may also bridge the gap between biodiversity and speciation studies. Many explanations have been proposed for the existence of biodiversity hotspots. Strangely, speciation has often been ruled out [72] as a mechanism of biodiversity, mainly for want of geographical features that can impose long-term isolation. With MIM cycles, the stringent requirement is relaxed and many geographical features could conceivably drive speciation. In the IWP, because the sea floor in the Indomalayan region has been relatively high, many shallow barriers have existed throughout the region [73].
When the global sea level began to decrease and fluctuate around that lower level 25 Myrs ago [17], many parts of the Indo-Malayan coasts may have experienced cycles of isolation and admixture. Indeed, as Renema et al. have pointed out, species diversity in the Indo-Malesia started to increase during the Miocene [23,74].

The MIM model may be applicable to other high-diversity spots as well. In the same time frame as mangrove speciation on the Indo-Malayan coasts, islands of the Aegean Archipelago in the Mediterranean may have been repeatedly connected and disconnected due to sea-level changes. Thus, the radiation of annual plants in the genus *Nigella* across this archipelago [75] could have also been driven by a mechanism like the MIM model. Similarly, MIM cycles may have driven: the extraordinary diversity of cichlid fish in Lake Victoria, which has experienced repeated rises and falls of water level [76]; diverse flora in neo-tropical rain forests subject to periods of cooler and drier climates driven by cyclical glacial events [77]; and avian diversification in the neo-tropics where fragmentation and reconnection of high-elevation habitats occurred during the late Pleistocene [78].

When diverging populations become fully fledged species, migration in the next M phase would be equivalent to range expansion. If speciation occurs after each isolation phase, there can be as many as $2^n$ species after $n$ cycles [58]. In that sense, the migration phase in the MIM cycles would play an added role in the evolution of biodiversity. More generally, isolation may create $i$ fragmented populations. If speciation is achieved after $j$ cycles, then the number of species after $n$ cycles would be $[i]^n/j$. In other words, the number of species after $n$ cycles can potentially be $m^n$ where $m = i^{n}/j > 1$. In the special case of $i = 2$ and $j = 1$, $m = 2^n$. Centers of high biodiversity are fascinating phenomena with many possible causes [20–22,79,80]. We suggest that efficient speciation mechanisms like MIM cycles may play a role.

**MATERIALS AND METHODS**

**Geographical distribution of mangrove species in IWP**

The geographical distribution of each of the nine mangrove genera in the IWP (*Kandelia, Aegiceras, Lumnitzera, Ceriops, Xylocarpus, Bruguiera, Avicennia, Rhizophora and Sonneratia*) was compiled from *World Mangrove ID* [33]. Species distribution ranges of *Ceriops* were updated according to Tsai et al. [81]. The distributions of *Rhizophora* and *Sonneratia* in China were updated from the field survey data of Wang and Chen [82].

**Scanning the genome for speciation islands**

To identify genomic regions highly divergent between *R. mucronata* and *R. stylosa*, we performed a genome-wide divergence scan using absolute measures of differentiation. Re-sequencing data of one *R. mucronata* individual from Ranong, Thailand, and one *R. stylosa* individual from Hainan, China, were generated using Illumina HiSeq 2000 platform. Reads were mapped to the *R. apiculata* reference genome using the BWA software [83]. Heterozygous sites were called using the GATK pipeline [84]. We used sliding windows to scan divergence levels between the two species. We set the window size to 50 Kb and step size to 25 Kb. Windows with fewer than 10 000 sequenced sites were discarded. The divergence level of each retained window was calculated as the number of differentiated sites divided by the number of sequenced sites. Divergent sites were defined as loci homozygous within each species but different between taxa.

**Sampling, sequencing and mapping**

We collected leaf material from populations of five mangrove species from 15 stands in the three regions as shown in Fig. 1a and Supplementary Table 1, available as Supplementary Data at NSR online. For each species, at least one stand was sampled in each region and 19–100 individuals were collected from each stand. Intervals between sampled individuals were at least 5 meters. Sequencing protocols were as described in our earlier work [32]. An equal amount of leaf material from each individual in every stand was mixed before DNA extraction. Based on sequences from cDNA libraries of the species, we designed primers for over 150 loci for each species. We succeeded in amplifying approximately 70 genes per species (Supplementary Table 2, available as Supplementary Data at NSR online) and sequenced them using the Illumina GA-II/HiSeq 2000 platform. The short reads sequence data were deposited in NCBI, BioProject: PRJNA303892. We mapped short reads to references using MAQ [85] with main parameters -m 0.002 and -e 200 and the parameter -q 30 to filter low-quality reads. To reduce sequencing errors, we ignored bases that were (i) located in the first 11 bp or the last 7 bp of the mapped reads, (ii) with base quality less than 22 and (iii) with minimum coverage less than 100.
Putative single nucleotide polymorphisms (SNPs) were called if the minor allele frequency was > 0.01.

**Haplotype inference**

Using the linkage information for SNPs in each pair of short reads, we estimated haplotypes and their frequencies using an expectation-maximization algorithm [86,87]. We divided the genes into two or more segments if the distance between two SNPs was longer than the length covered by paired reads. To validate the estimated haplotype phases, we sequenced 360 alleles in eight populations using the Sanger method (Supplementary Dataset 1, available as Supplementary Data at NSR online). Haplotypes and their frequencies estimated using these two approaches were very similar. Short reads were informative for our haplotype analyses thanks to large sample sizes. We constructed haplotype networks for each gene segment based on the inferred haplotypes.

**Estimating nucleotide diversity and population structure**

Using the obtained haplotype profiles, we estimated nucleotide diversity ($\pi$, the average number of nucleotide mismatches per site between two sequences [88]) within a stand/population ($\pi_s$), region ($\pi_R$, with all areas weighted equally) and species ($\pi_T$, with all three regions weighted equally) for each haplotype segment. We employed F-statistics at different levels to measure population differentiation ($F_{ST} = 1 - \pi_s/\pi_T$) [49].

After carefully reviewing haplotype networks for the five species, we observed that haplotypes of many genes could be clustered into two distinct clades corresponding to the samples from the East and West Indo-Malayan regions. We therefore clustered the haplotypes into two clades using complete linkage method hierarchical clustering analysis [86]. We included segments with more than two SNPs, or two SNPs and two haplotypes. We calculated frequencies of haplotype clades in both regions.

**Demographic models and parameter estimation**

We used a maximum-likelihood method to estimate effective population size ($N_e$) and migration rate ($m$) for the SIM and MIM models. The SIM model requires an additional parameter: the isolation time imposed by the geographical isolation, as depicted in Fig. 3a. The time elements in the MIM model were defined by the geological records of sea-level changes. The mutation rate was inferred from exome/transcriptome data with the fossil records as described in the Table 1 legend.

The number of divergent nucleotides between two sequences sampled from the same populations was denoted as $D_w$, while differentiation between populations was denoted as $D_b$ ($w$ stands for within population and $b$ for between populations). The log-likelihood function can be constructed as follows:

$$
\log(L) = \log \left( \prod_x P(D_b = x)^{f(D_b = x)} \times \prod_y P(D_w = y)^{f(D_w = y)} \right)
= \sum_x f(D_b = x) \log(P(D_b = x)) + \sum_y f(D_w = y) \log(P(D_w = y)).
$$

(1)
Simulations of DNA-sequence evolution

We used ms [89] to simulate sequence evolution under the MIM scheme for 2000 replicates in each species. Parameter values used in simulations are listed in Table 2. Six statistics were obtained from the simulated sequences for each species: \( D_{\text{within}} \) (average divergence within region), \( D_{\text{between}} \) (average divergence between regions), \( D_{\text{clade}} \) (differentiation among the most recent common ancestors of each clade), \( D_{\max} \) (differences between two most divergent haplotypes), \( F_{\text{ST}} \) (total number of SNPs) and \( F_{\text{PSM}} \) among regions. The simulated distributions of the six statistics are comparable to values observed from data (Supplementary Figs 8–12, available as Supplementary Data at NSR online).

We also simulated 2000 replicates under the SIM evolution scheme using the parameters listed in Table 2 for calculating \( D_{\max} \). We calculated the standard deviation of \( D_{\max} \) among genes in each relicate derived from the SIM or MIM model. The distributions of the 2000 standard deviations from the two models are depicted in the insets of Fig. 4a–c.

To test whether the MIM model fits observed data better than the SIM model, we obtained MLE of the two models for sequences simulated under the SIM model. We calculated the differences in the likelihood values (Diff = log-likelihood of the MIM model – log-likelihood of SIM model) for each of the 2000 repetitions. For \( R. \) apiculata, \( S. \) alba, \( A. \) marina and \( A. \) coriinulatum, the Diff value of the real data is larger than all the Diff values of the simulated sequences. Hence, the probability that the SIM model fits data better than the MIM model is less than 0.001. For \( C. \) tagal, the probability is 0.33. As discussed in the main text, the unusually low genetic diversity of \( C. \) tagal makes it powerless to compare models.

Estimating effective population-size change using whole-genome sequence data

To estimate past effective population size, we used the PSMC [51]. We used the whole-genome sequence data from six individuals (data deposited in NCBI, BioProject: PRJNA298659). \( A. \) marina and \( S. \) alba individuals were from the Gulf of Thailand and the West Coast; \( R. \) apiculata samples were from Sanya and Wenchang. We mapped the re-sequencing data generated by the Illumina HiSeq 2000 platform to the corresponding draft genomes ([26] and He et al., unpublished data) using BWA [83]. The parameters of PSMC estimation were: -N25 -t15 -r5 -p "4+25*2+4+6". Generation time was set to 20 years. The mutation rate for each species is given in Table 2.

We also produced simulated sequence data for PSMC analysis (see Fig. 4d and Supplementary Figs 13–15, available as Supplementary Data at NSR online). The simulated sequences were generated by msHOT [90] with the following parameters: mutation rate (\( \mu \)) set as 0.5, 1.0, 2.0 \( \times 10^{-7} / \text{site/year} \), migration rate (\( m \)) as \( 1 \times 10^{-4}, 5 \times 10^{-4}, 10 \times 10^{-4} \) per generation, population size (\( N \)) as 100, 500, 1000 and 5000. Each simulated genome contained 500 loci and the length of each gene was set to 200 Kb. The recombination rate was set to \( 1 \times 10^{-9} / \text{site/generation} \).

SUPPLEMENTARY DATA

Supplementary data are available at NSR online.

ACKNOWLEDGEMENTS

We thank David Jablonski, Roger Butlin, Richard Abbott, Trevor Price, Loren Rieseberg, Nick Barton, Patrik Nosil, James Mallet, Daven Presgraves, Dolph Schluter, Fangliang He, Chuck Canon and Xiongwei He for insightful comments that improved the manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (91731301 and 31600182); the National Key Research and Development Plan (2017FY100705); the 985 Project (33000-18831107) and the Fundamental Research Funds for the Central Universities (17lggy99).

AUTHOR CONTRIBUTIONS

C.-L.W. and S.S. conceived the project and wrote the manuscript; Z.H., X.L. and M.Y. designed the experiments, analysed data and prepared the manuscript; Z.H. and C.-L.W. constructed demographic models; S.S., X.L., C.Z. and N.D. contributed to sample collections. X.L. and X.W. performed experiments; N.D. contributed to manuscript modifications.

ACCESSION CODES

Illumina reads are available at the Short Read Archive under the NCBI BioProject PRJNA303892 and PRJNA298659.

REFERENCES

15. Cruickshank TE and Hahn MW. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 2014; 23: 3133–57.
Allo-parapatric speciation goes offshore

A basic conundrum in speciation is that isolation is required for differentiation, but dispersal is required to produce populations in different places, which can then diverge [1]. These conflicting pressures seem to make the optimal conditions for speciation to be those in which barriers appear and disappear on the right timescale. If barriers break down too quickly, divergence will be incomplete and populations will merge; if they appear too slowly, few species will form. The timescale of barrier fluctuations is known in some cases, such as from sea-level fluctuations. In this novel application of genomics and computer modeling, He et al. connect paleoclimates to recent mangrove speciation and differentiation. They demonstrate that speciation has occurred despite periods of contact. By showing that genetic differentiation along the genome is too variable to be accounted for by a simple allopatric model, they confirm that gene flow between the diverging species must have happened, with timescales for contact on the order of 100,000 years. Further, they use a novel approach to estimate the minimum time for complete gene flow to be cut off if divergence were to happen entirely in allopatry, through comparisons of subspecies, and show that this is too long to be a feasible means of creating species diversity.

The paper combines cutting-edge genomic analysis with an impressive dataset to draw strong conclusions about times needed to create enough divergence that populations will not collapse back into one upon coming into contact. It also leads to increased understanding of mechanisms producing the extraordinary marine diversity in the region. Finally, the results set the stage for asking whether contact between divergent forms, such as in hybrid zones, could actually accelerate speciation rather than retard it, such as by promoting chromosomal inversions [2] or driving pre-zygotic reinforcement once some post-zygotic incompatibilities have developed [3].

Trevor Price
Department of Ecology and Evolution,
University of Chicago, USA
Recommender of NSR
E-mail: pricet@uchicago.edu

REFERENCES

National Science Review
6: 289, 2019
doi: 10.1093/nsr/nwy110
Advance access publication 15 October 2018

A new model of speciation

In the first paragraph of ‘On the origin of species’, Darwin [1] refers to speciation as that ‘mystery of mysteries’. Although the means by which new species arise is no longer mysterious, some puzzles remain. One such puzzle, which is addressed in a recent paper from Chung-I Wu’s group on mangrove speciation [2], concerns the popular allopatric speciation model. In this model, which is widely viewed as the most common form of speciation, geographically isolated populations diverge to the point at which successful reproduction between them is no longer possible. However, there are too few geographical features that can offer stable isolation over long enough periods of time to account for the number of coexisting species that we see in the world today. Likewise, population genomic data reveal that a surprisingly high fraction of species that have been analyzed so far exhibit footprints of hybridization in their genomes. While models of speciation with gene flow (e.g. parapatric and sympatric speciation) can account for such a signature of hybridization, the widespread presence of postzygotic isolating barriers, even in taxa that currently hybridize, implies that periods of allopatry during divergence are common as well.

The mixing–isolation–mixing model (MIM) proposed by Ziwen et al. [2] offers a possible resolution to this conundrum. As suggested by the name, the model suggests that speciation often involves repeated cycles of admixture followed by geographical isolation, until full isolation is ultimately achieved. Such a model is difficult to prove based on population genomic data alone, so the authors have compiled a comprehensive array of genomic, geohistorical and theoretical data to support their proposed model in the mangrove system.

As with most concepts in evolutionary biology, the MIM model is not completely new. It bears some resemblance, for
example, to Ehrendorfer’s differentiation–hybridization cycles [3] and Rattenbury’s cyclic hybridization model [4]. However, Ehrendorfer felt that hybridization would break down existing differentiation and that a new round of differentiation would then emerge from the hybrid complex. Thus, his views are arguably more similar to Seehauzen’s proposal [5], that adaptive radiations might spring from the wide diversity generated by hybrid swarms, than to MIM. Rattenbury placed greater emphasis on the role of hybridization in re-acquiring key adaptations that would allow them to survive climate oscillations than on the importance of allopatric periods in generating postzygotic reproductive isolation. More generally, there has been widespread recognition since the 1970s that many taxa have experienced repeated episodes of allopatry and parapatry due to cyclic variation in climate. However, the emphasis of this work was more on phylogeography and hybrid zone dynamics than on a particular speciation mechanism such as MIM. It is also important to note that none of these earlier authors, to my knowledge, viewed hybridization–differentiation cycles as a solution to the conundrum of too little allopatry to account for patterns of species richness. Thus, I view the MIM proposal as a novel and important contribution to the field of speciation.

A mixing–isolation–mixing model of speciation can potentially explain hotspots of species diversity

From comparative genomic analyses between species and population genetic analyses within species, He et al. [1] provide convincing evidence that mangrove divergence across the Malacca Straight, separating the Indian and Pacific Oceans at the Malay Peninsula, Southeast Asia, fits a model of speciation occurring over a time span that includes cycles of isolation (allopatry) interspersed by periods of mixing (gene flow). Key to this demonstration was the known biogeography of the region, and the fact that rises and falls in sea level during the period of divergence and speciation have repeatedly opened and closed the Malacca Straight to migration between the Indian and Pacific Oceans on either side of the Strait. Without this knowledge, it would have simply been concluded that speciation had occurred with gene flow, thus failing to recognize the occurrence of allopatric phases and their likely importance in the build-up of reproductive incompatibilities between taxa [2].

More generally, He et al. [1] show that their proposed mixing–isolation–mixing (MIM) model of speciation can result in a large increase in the number of species that form in areas where impermanent geographical barriers allow recurrent cycles of isolation and mixing to occur. This is the case if speciation occurs after each of several isolation phases, which may explain why hotspots of species diversity occur in some regions where such impermanent geographical barriers (rather than longstanding ones) are present. They further suggest that the MIM mechanism of speciation may explain the occurrence of rapid species radiations in some taxonomic groups. Possible examples include the radiation of cichlid fish in Lake Victoria, East Africa, where repeated cycles in the rise and fall of water levels have occurred during the period that covers this radiation, and the radiation of the plant genus Nigella across the Aegean Archipelago in the Mediterranean [3]. The Aegean Archipelago originated ∼5 Ma, with some islands repeatedly connected and disconnected during the Pleistocene due to falls and rises of sea levels in glacial and interglacial periods, respectively. The terrestrial flora and fauna of this and other continental shelf island archipelagos provide ideal material for further testing of the MIM mechanism.

The MIM mechanism of speciation is of further relevance to an understanding of hybrid zone formation between taxa that are not fully reproductively isolated from each other. Formation of hybrid zones may occur in either of two ways, by primary intergradation where divergence occurs in the face of gene flow or by secondary contact between populations that diverged in allopatry [4,5]. However, hybrid zones formed in either way exhibit very similar patterns of genetic and phenotypic variation at equilibrium [4,6,7], and, consequently, their exact mechanism of origin cannot be determined in the absence of a historical record. Attempts have been made to reconstruct from genomic data the way in which some hybrid zones have

REFERENCES

doi: 10.1093/nsr/nwy111
Advance access publication 8 October 2018
originated, leading to the conclusion that an origin in the face of continuous gene flow (primary intergradation) is a satisfactory explanation in some instances (see, for example, Filatov et al. [8]). However, without a historical record, it is not possible to rule out the occurrence of allopatric phases, which may have been key to the development of prezygotic and postzygotic isolating barriers that reduce gene flow between parental types [9]. The convincing demonstration of the MIM model to explain mangrove speciation across the Indo-Pacific Barrier at the Malay Peninsula sets the scene for future studies of speciation (and the origin of hybrid zones) in all regions where an underlying historical record of biogeography is available or can be obtained.

Richard J. Abbott
School of Biology, University of St Andrews, UK
Reviewer of NSR
E-mail: rja@st-andrews.ac.uk

REFERENCES

National Science Review
6: 290–291, 2019
doi: 10.1093/nsr/nwy112
Advance access publication 8 October 2018

Is speciation driven by cycles of mixing and isolation?

He et al. [1] describe genetic divergence between mangrove species that are distributed across the Straits of Malacca. Divergence varies much more across the genome than in a simple null model of strict allopatry. Such heterogeneity (often termed ‘genomic islands of speciation’) can be generated by population structure: populations that exchange genes at a low rate will show wide variability in divergence, depending on when migration events occurred on the ancestral lineages that trace back from each section of genome. He et al. [1] also used the pairwise sequentially Markov coalescent to estimate that the species’ effective population size increases back into the past. Again, this can be explained by population subdivision: lineages sampled from the same place either quickly coalesce, or wander across the species’ range as they trace back into the past, reducing the rate of coalescence, and therefore increasing the species-wide effective size. These observations are consistent with intermittent isolation and mixing caused by changes in sea level. However, as He et al. [1] acknowledge, they may be produced by many other processes: for example, selection or variation in recombination rate [2]. Thus, the best evidence for intermittent mixing is geological.

In their abstract, He et al. [1] argue that ‘speciation... is driven by cycles of isolation and gene flow’, and that ‘the mixing–isolation–mixing (MIM) mechanism of speciation is... efficient, potentially yielding mn (m>1) species after n cycles’. Both claims can be contested.

First, there is no evidence as to the causes of speciation. In this mangrove example, we have good geological evidence for intermittent contact and we see high species diversity. However, to show a connection between the two, one would need to establish that the actual reproductive isolation and ecological divergence were caused (or at least facilitated) by intermittent gene flow. This is extremely challenging. Alternatively, a comparative study might show a significant association between intermittent geographic barriers and species diversity. This seems more feasible, but nevertheless would require very many replicate species groups.

Second, He et al. [1] argue that MIM gives an efficient mechanism of speciation, because each cycle multiplies the number of species by some factor, leading to an exponential growth in species numbers. Yet, any speciation mechanism can potentially lead to exponential growth if one assumes that each species can split into multiple species at some rate (of course, in reality, species numbers are limited by available locations and niches). More fundamentally, this argument takes ‘speciation’ to be a discrete event. In fact, it is a process by which multiple differences evolve independently, and then gradually come together to give distinct and incompatible populations that we recognize as species [3]. Multiple episodes of secondary contact (i.e. of MIM) or, less drastically, expansions and contractions of local populations, can bring together differences that evolved independently, producing hybrid zones and ultimately full species...
This process may play an important role in speciation; yet, even with an abundance of genomic data, it is hard to see how to estimate its importance.

Nicholas H. Barton
Institute of Science and Technology Austria, Austria
Reviewer of NSR
E-mail: nick.barton@ist.ac.at

REFERENCES

National Science Review
6: 291–292, 2019
doi: 10.1093/nsr/nwy113
Advance access publication 8 October 2018