MOLECULAR ECOLOGY

Molecular Ecology (2013) doi: 10.1111/mec.12442

Restricted gene flow within and between rapidly diverging Neotropical plant species

YANN SURGET-GROBA* and KATHLEEN M. KAY†

*Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Menglun, Yunnan 666303, China, †Ecology and Evolutionary Biology, University of California, Santa Cruz, 1156 High St., Santa Cruz, CA 95064, USA

Abstract

Speciation involves the evolution of traits and genetic differences that contribute to reproductive isolation and the cessation of gene flow, and studying closely related species and divergent populations gives insight into how these phenomena proceed. Here, we document patterns of gene flow within and between two members of a rapid Neotropical species radiation, *Costus pulverulentus* and *Costus scaber* (Costaceae). These species co-occur in the tropical rainforest and share pollinators, but are reproductively isolated by a series of prezygotic barriers, some of which show evidence of reinforcement at sympatric sites. Here, we genotype microsatellite markers in plants from eight sites that span the geographical range of both species, including four sympatric sites. We also genotype putative hybrids found at two sympatric sites. We find high levels of genetic isolation among populations within each species and low but detectable levels of introgression between species at sympatric sites. Putative hybrids identified by morphology are consistent with F1 or more advanced hybrids. Our results highlight the effectiveness of prezygotic isolating mechanisms at maintaining species boundaries in young radiations and provide empirical data on levels of gene flow consistent with reinforcement.

Keywords: gene flow, hybridization, speciation, tropical plants

Received 9 November 2012; revision received 2 July 2013; accepted 4 July 2013

Introduction

Species radiations depend on the evolution and maintenance of barriers to gene flow, yet many closely related species lack strong intrinsic forms of isolation and are able to form fertile hybrids when mated (Seehausen *et al.* 2008; Schemske 2010). For co-occurring species, strong prezygotic isolating barriers may prevent most hybridization (Mayr 1947, 1959; Ramsey *et al.* 2003; Kay 2006). In other cases, hybrids may be formed but exhibit low fitness in the wild, such that the backcrossing required for introgression is rare (Rundle & Whitlock 2001; Egan & Funk 2009; McBride & Singer 2010). At the other extreme, strong selection may maintain differences in key morphological and physiological traits in the face of extensive hybridization and introgression (Barton & Bengtsson 1986; Grant 1993; Wu 2001; Turner *et al.* 2005; Nadeau *et al.* 2012; Renaut *et al.* 2012).

Radiations of species with weak intrinsic isolation may be especially likely in situations in which allopatric divergence among populations happens over short timescales and small geographical scales. The same geographical features may contribute to isolation very differently across a range of organisms and ecological settings. Sedentary organisms and organisms with limited dispersal can become geographically isolated more easily (Coyne & Orr 2004). Hence, the time to speciation and the geographical area needed for allopatric speciation are inversely correlated with typical magnitudes of F_{ST} across major groups of organisms, with snails and plants having much higher probabilities of speciation in a given area than widely dispersing animals (Kisel & Barraclough 2010). Likewise, geographical regions with little temporal environmental variance may give rise to species with narrow physiological

Correspondence: Kathleen M. Kay, Fax: 831 459 5353;
E-mail: kmkay@ucsc.edu

tolerances and limited long-range dispersal (Janzen 1967). When combined with divergent selection across the geographical range, this can lead to high rates of speciation (Schemske 2002). To understand the origin and maintenance of species radiations, we need to assess patterns of gene flow among populations within species at the initial stages of divergence and quantify any gene flow between recently diverged species experiencing sympatric contact.

Neotropical plant radiations provide a perfect system for examining population divergence, sympatric coexistence and potential gene flow among close relatives. The Neotropical forests are exceptionally diverse, and much of that diversity is relatively recent, especially among herbs, shrubs and epiphytes (Gentry 1982; Hoorn *et al.* 2010). Moreover, it is common to find many congeners co-occurring in the same forest sites. Many of these species that co-occur are known to be interfertile in artificial crosses and yet maintain clear morphological differences in nature (e.g. Ippolito *et al.* 2004; Smith & Baum 2006; Kay & Schemske 2008). Nevertheless, for tropical plants, we know very little about natural hybridization and gene flow between species or about levels of gene flow among geographically disparate populations within species.

Here, we investigate patterns of intraspecific and interspecific gene flow among populations of two recently derived species of Neotropical spiral gingers. The genus *Costus* comprises approximately 60 species of perennial monocots in the Neotropics and has undergone one of the fastest known plant radiations since their colonization of Central America approximately 1.1–5.4 Ma from Africa (Kay *et al.* 2005). Throughout the Neotropics, multiple species of *Costus* commonly occur sympatrically, and species are widely interfertile in greenhouse crosses (Kay & Schemske 2008). *Costus pulverulentus* and *Costus scaber* are closely related species that have partially overlapping ranges, share hummingbird pollinators and flower at the same time (Kay & Schemske 2003; Kay 2006). They are both part of an unresolved clade together with other closely related Central American species (*Costus spicatus*, *C.* aff. *wilsonii*, *Costus woodsonii*) that diverged between 0.3 and 1.6 Ma (Kay *et al.* 2005). Although fertile hybrids can be made experimentally, there is nearly complete prezygotic isolation in the field due to differences in habitat affinity, floral mechanical isolation and pollen–pistil incompatibility (Kay 2006), and obvious morphological hybrids are rare. Moreover, the pollen–pistil incompatibilities have evolved only between locally sympatric populations, consistent with the process of reinforcement (Kay & Schemske 2008).

By genotyping a set of rapidly evolving microsatellites, we sought to answer several questions about gene flow both within and between these species. First, we asked how isolated populations within a species are from each other across a range of geographical distances using summary genetic distances and multivariate analysis. We predicted that the rapid species radiation observed within this genus would be associated with restricted gene flow among populations. We analysed all sampled populations of both species concurrently in order to compare levels of isolation within species to those found between species. Next, we asked whether there is evidence for hybridization and introgression between species. For close relatives, postspeciation gene flow is notoriously difficult to distinguish from shared ancestral polymorphism (Nielsen & Wakeley 2001). The difference can be modelled (e.g. Kuhner 2006; Hey 2010), but the models make many simplifying assumptions about population structure and stability that are unlikely to be met. Alternatively, genotypes may be sampled in a geographically structured way such that there are explicit hypotheses about what geographical patterns would indicate recent interspecific gene flow (e.g. Kane *et al.* 2009; Palma-Silva *et al.* 2011). Here, we asked whether there is admixture or depressed genetic distance between sympatric populations compared with allopatric populations and whether phenotypically intermediate individuals from sympatric sites are evidence for hybridization and introgression in this putatively reinforced system.

Materials and methods

Sampling

Sympatric sites, from which both species were collected, included La Selva Biological Station (LS) and Sirena Biological Station (OSA) in Costa Rica, and Barro Colorado Island Nature Monument (BCI) and along Pipeline Road (PLR) in Panama (Fig. 1; Table 1). Allopatric populations of *C. scaber* were collected along the Rio Moile (MOI) and in Madidi National Park (MAD) in Bolivia, and allopatric populations of *C. pulverulentus* were collected near the town of Lacanja (LAC) and near the Mayan ruins of Bonampak (BON) in Chiapas, Mexico (Fig. 1; Table 1 and Table S1, Supporting information). La Selva (10.428637°N 84.013137°W) is a 1536-ha reserve in the Atlantic lowlands of Costa Rica that shares a boundary with the extensive Braulio Carrillo National Park. It consists of mature tropical wet forest as well as regenerating agricultural lands (McDade & Hartshorn 1994). OSA (8.473723°N 83.589037°W) is located along the Pacific coast of Costa Rica in the expansive Area Conservación de Osa, and the surroundings consist of mature tropical wet forest and regenerating agricultural lands. BCI (9.155042°N

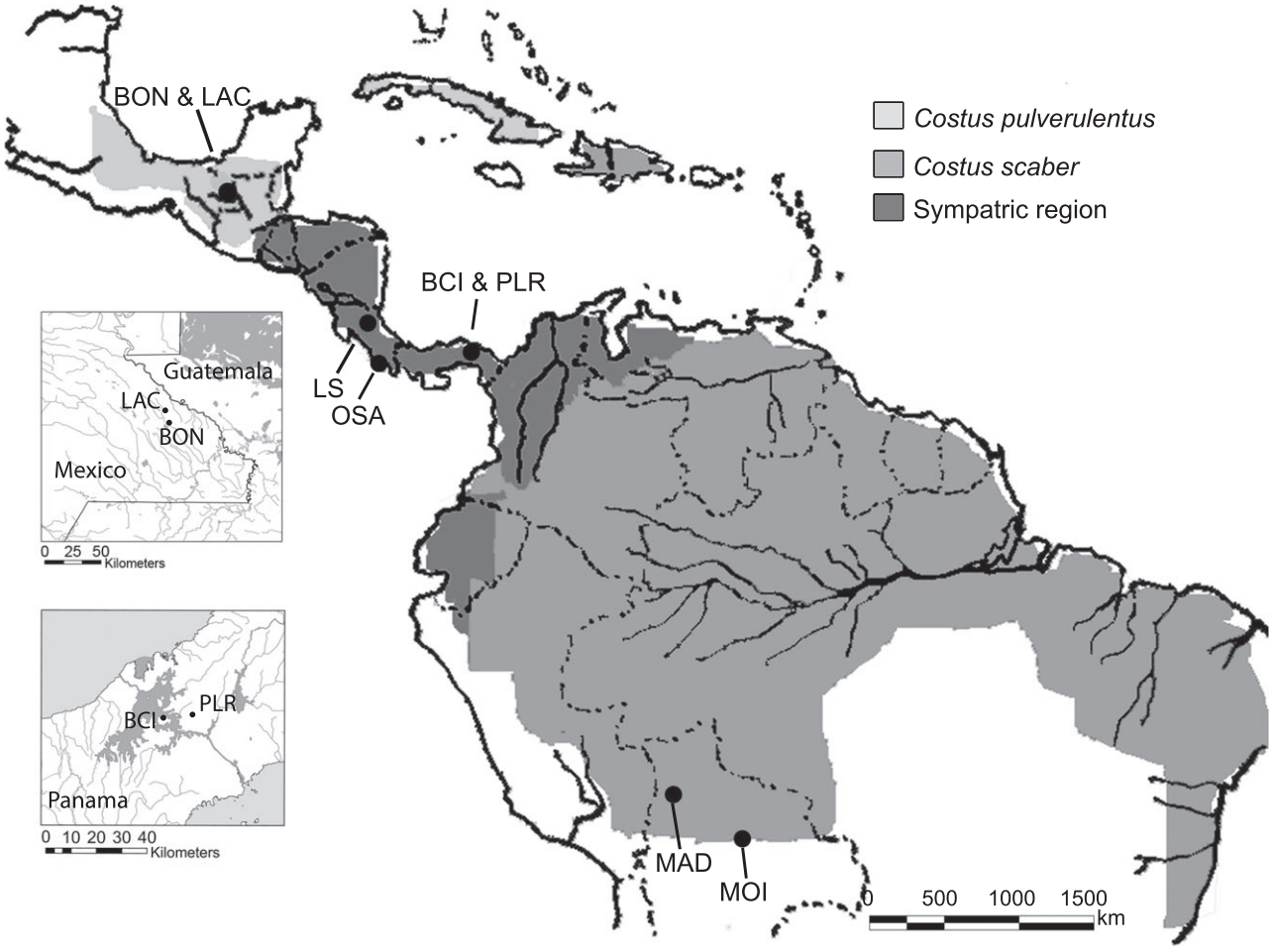

Fig. 1 Distribution map of *Costus pulverulentus* and *Costus scaber*. The detailed insets are areas in Mexico and Panama where the sampled populations are too close to be distinguished on the general map.

Table 1 Details of the populations sampled. Code: the first letter of the population code corresponds to the species (p: *Costus pulverulentus*, s: *C. scaber*); N: number of individuals analysed; NA: average number of alleles per locus; H_O: average observed heterozygosity over loci; H_E: average heterozygosity over loci; HW: number of loci deviating from HW expectations after Bonferroni correction; PA: number of private alleles

Country	Site	Code	*N*	NA (minimum–maximum)	H_O	H_E	HW	PA
Mexico	Bonampak	pBON	11	2.69 (2–6)	0.37	0.38	0	4
	Lacanja	pLAC	11	2.89 (2–5)	0.55	0.52	0	3
Costa Rica	Sirena Biological Station	pOSA	20	3.58 (2–9)	0.36	0.39	0	3
		sOSA	19	3.42 (2–8)	0.43	0.45	0	8
	La Selva Biological Station	pLS	27	3.91 (2–13)	0.35	0.44	2	6
		sLS	26	3.53 (2–13)	0.26	0.33	0	5
Panama	Barro Colorado Island	pBCI	25	2.44 (2–4)	0.10	0.25	4	2
		sBCI	25	2.71 (2–4)	0.53	0.44	0	0
	Pipeline Road	pPLR	22	2.11 (2–3)	0.09	0.25	4	2
		sPLR	24	3.29 (2–5)	0.52	0.43	0	2
Bolivia	Madidi National Park	sMAD	6	3.20 (2–7)	0.34	0.52	0	1
	Rio Moile	sMOI	15	2.75 (2–5)	0.29	0.34	0	2

79.847996°W) is a 16-km^2 island located in Gatun Lake in the Panama Canal and consisting of mature tropical moist forest (Croat 1978). PLR (9.171599°N 79.753202°W) is in a conservation region immediately adjacent to the Panama Canal and is dominated by secondary tropical moist forest. The MOI site (17.493801°S 63.810245°W) is located on the border of the 430 000 ha Parque Nacional y Área de Uso Múltiple Amboró in Bolivia, in fragments of seasonal moist forest. MAD (14.425403°S 67.920282°W) consists of mature seasonal moist forest. LAC (16.798721°N 91.105660°W) and BON (16.706024°N 91.064162°W) are nearby sites in a mixture of primary and secondary evergreen tropical moist forest in the Lacandon Jungle. Voucher information for these populations can be found in Kay *et al.* (2005).

We collected young leaf tissue from 6–27 flowering individuals per population (average, 18.25; median, 21) from four sympatric sites and two allopatric sites for each species. In forested regions, *Costus* grow at low density, sometimes with hundreds of metres between flowering individuals. We typically sampled all individuals encountered. Sampling was restricted to flowering individuals because these species, along with other co-occurring *Costus*, cannot reliably be distinguished when vegetative. In addition, we sampled tissue from all putative hybrids encountered, which were identified by their intermediate morphology and similarity to greenhouse-created hybrids. Four putative hybrids were collected at Osa and three at PLR. Of the four collected at Osa, three appeared to be hybrids between the study species and one appeared to be a hybrid between *C. pulverulentus* and *Costus lima*, another sympatric species.

Genetic analyses

For each individual, we extracted genomic DNA using the DNeasy Plant Mini Kit (Qiagen) and genotyped 15 microsatellite loci developed from EST sequences. Details on microsatellite development and PCR conditions are described in Molecular Ecology Resources Primer Development Consortium *et al.* (2012). Amplified products were sized on an ABI 3730×l Genetic Analyzer, and genotypes were scored using GeneMapper v4.0.

We checked for deviations from Hardy–Weinberg equilibrium (HWE), calculated observed and expected heterozygosity, quantified the distribution of genetic variance within and between species using AMOVA and estimated levels of population differentiation with pairwise standardized F_{ST} (G'_{ST}, Hedrick 2005) in Arlequin v3.5.1.2 (Excoffier & Lischer 2010) and RecodeData v0.1 (Hedrick 2005; for the G'_{ST} calculations). Data were treated as standard data, because a few loci did not follow a stepwise mutation model. We used 1000 permutations and a Bonferroni correction to test the significance of the F_{ST} values. We tested for isolation by distance both within and between species by comparing a matrix of geographical distances (log corrected) to a matrix of Slatkin's linearized F_{ST} $\{F_{ST}/(1-F_{ST})$; (Rousset 1997)$\}$, using a Mantel test with 1000 permutations to test significance. We predicted that populations would show isolation by distance in interspecific comparisons if there had been substantial introgression in sympatric regions.

We used the Bayesian clustering method implemented in STRUCTURE v2.3.2 (Pritchard *et al.* 2000) to identify distinct clusters in our data, assess likely routes of intraspecific gene flow and look for evidence of interspecific admixture of sympatric populations. We identified the most likely number of clusters (K) by running the analysis with specified priors for K ranging from one to the number of sampled populations of both species plus two sets of hybrids (14 total). For each K value, the analysis was run 20 times using the admixture model, with a burn-in of 100 000 steps for a total run length of 500 000 steps. The optimal K was inferred following the method outlined in Evanno *et al.* (2005), identifying the number of clusters corresponding to the highest rates of change in the log probability of data between successive K values and the guidelines from the STRUCTURE manual. When including distant sites in an overall analysis, homoplasy can distort inferences of admixture in sympatry. Therefore, we also applied the STRUCTURE clustering analysis separately for each of the four sympatric sites (with 10 replicate runs, a burn-in of 100 000 steps and a total run length of 500 000 steps). We fixed $K = 2$ and estimated the rate of admixture by the proportion of the genome of the individuals of each species assigned to the other species.

Putative hybrids identified by intermediate morphology were further analysed with the Bayesian clustering approach implemented in NewHybrids (Anderson & Thompson 2002) to estimate the posterior probability that they represent F1 hybrids, F2 hybrids or first generation backcrosses to either parent. We used the individuals with more than 99% of their genome assigned to either species in the STRUCTURE analysis as a 'pure' parental reference for the analysis. We conducted these analyses separately for the plants from PLR and OSA, using a burn-in of 50 000 steps followed by a run length of 300 000 steps for each analysis.

We then used a discriminant analysis of principal components (DAPC, Jombart *et al.* 2010) to describe the global structure of populations using multivariate analysis. The microsatellite data were subjected to a PCA using the adegenet (Jombart 2008) and ade4 (Thioulouse *et al.* 1997) packages in R (R Core Development Team

2011). We then selected the components representing more than 95% of the variation of the initial data set and subjected these to a linear discriminant analysis using the MASS package (Venables & Ripley 2002) in R, using the population of origin as the grouping factor. This allowed us to test whether individuals can be confidently assigned to their population of origin, indicating high differentiation and low levels of gene flow, or not, indicating low differentiation and high levels of gene flow.

Results

Excluding the three populations with low sample size (<15, see Table 1), the genetic diversity (expected heterozygosity) varied from 0.25 for the populations of *C. pulverulentus* from Panama to 0.45 for *C. scaber* in Sirena, Costa Rica (Table 1). The populations of *C. pulverulentus* in Panama (BCI and PLR) that have been observed to set seed autogamously in the greenhouse (K. M. Kay, unpublished data) have the lowest diversity and show significant deviation from HWE at four loci and are fixed at six other loci, a pattern commonly observed in selfing populations. The number of alleles is modest with an average of 3.04 alleles per locus per population. The numbers of alleles in the Costa Rican populations are higher than those in the other populations.

We found strong genetic structure at both the population and species level, and most populations had one or more private alleles (Table 1). Highly significant AMOVA shows that 13.91% of the variance is found between species (Va = 0.348, $P < 0.001$), 38.27% among populations within species (Vb = 0.956, $P < 0.001$) and 47.79% within populations (Vc = 1.194, $P < 0.001$). All fixation indices are highly significant ($F_{ST} = 0.522$, $F_{SC} = 0.445$, $F_{CT} = 0.139$; $P < 0.001$ in all cases). Intraspecific pairwise G'_{ST} values were generally quite high, ranging from 0.064 to 0.693 in *C. pulverulentus*, and from 0.165 to 0.923 in *C. scaber* (Table S2, Supporting information). Interspecific values were also very high, but within the range of intraspecific values, ranging from 0.472 to 0.874. All pairwise comparisons were highly significant, except for *C. pulverulentus* between the close-by (approximately 11 km apart) populations of BCI and PLR ($G'_{ST} = 0.064$, $P = 0.047$, NS after Bonferroni correction).

The genetic distances were highly correlated with the geographical distance in *C. pulverulentus* ($r = 0.803$, $P = 0.006$), but the correlation was weaker and marginally significant in *C. scaber* ($r = 0.497$, $P = 0.076$; Fig. S1, Supporting information). It should be noted, however, that this test has limited power to detect significant relationships with only six sites. No correlation was observed when using only interspecific comparisons ($r = 0.175$, $P = 0.16$).

The STRUCTURE results showed two distinct peaks of delta*K*. The first one at $K = 5$ was the highest (delta$K = 411.3$), but at this peak, the likelihood ($L(K = 5) = -4318.3$) had not reached its final plateau and was much lower than that for the second peak at $K = 9$ ($L(K = 9) = -3606.1$; delta$K = 154.9$; Fig. S2, Supporting information), a situation very different from the simulations explored by Evanno *et al.* (2005). Therefore, we chose the *K* value with both a high likelihood and a peak of deltaK, and we considered that the genetic diversity of these species was best partitioned among nine clusters (Fig. 2).

Across the six sampled populations of *C. pulverulentus*, four clusters could be distinguished. The first one comprised the individuals of the two Mexican populations and the second one the individuals of the two Panamanian populations. Between these two clusters, almost no admixture was detected. The two populations from Costa Rica formed the remaining two clusters. Slightly more admixture was observed between these two clusters, especially in the LS population, for which several individuals had a significant portion of their genome assigned to the OSA Costa Rican cluster.

Four clusters could also be observed across the six sampled populations of *C. scaber*. The two populations from Bolivia formed the first cluster, the populations from Panama formed the second one, and the two populations from Costa Rica formed the last two. Here again, all clusters were very isolated and very little admixture was observed.

Four putative hybrids were sampled from OSA. Three of these had their genomes split almost equally between *C. pulverulentus* and *C. scaber* in the STRUCTURE analysis, and using NewHybrids, we estimated that they are F1 hybrids with 0.82, 0.80 and 0.95 posterior probability. The remaining hybrid had approximately half its genome assigned to the local *C. pulverulentus*, and the other half was a mix of several other clusters, consistent with our hypothesis that it represents a hybrid between *C. pulverulentus* and *C. lima*. We were unable to assign it to any hybrid class between *C. pulverulentus* and *C. scaber* with high posterior probability. In contrast, the three putative hybrids identified from PLR form a unique group in the STRUCTURE analysis. At four of the microsatellite loci, these plants had one or more alleles not found in either *C. pulverulentus* or *C. scaber* from PLR (one allele being absent from all the populations studied) and may include some ancestry of other sympatric *Costus* species that have not been sampled for this study. All three were assigned as backcrosses with *C. scaber*, with 0.78 posterior probability.

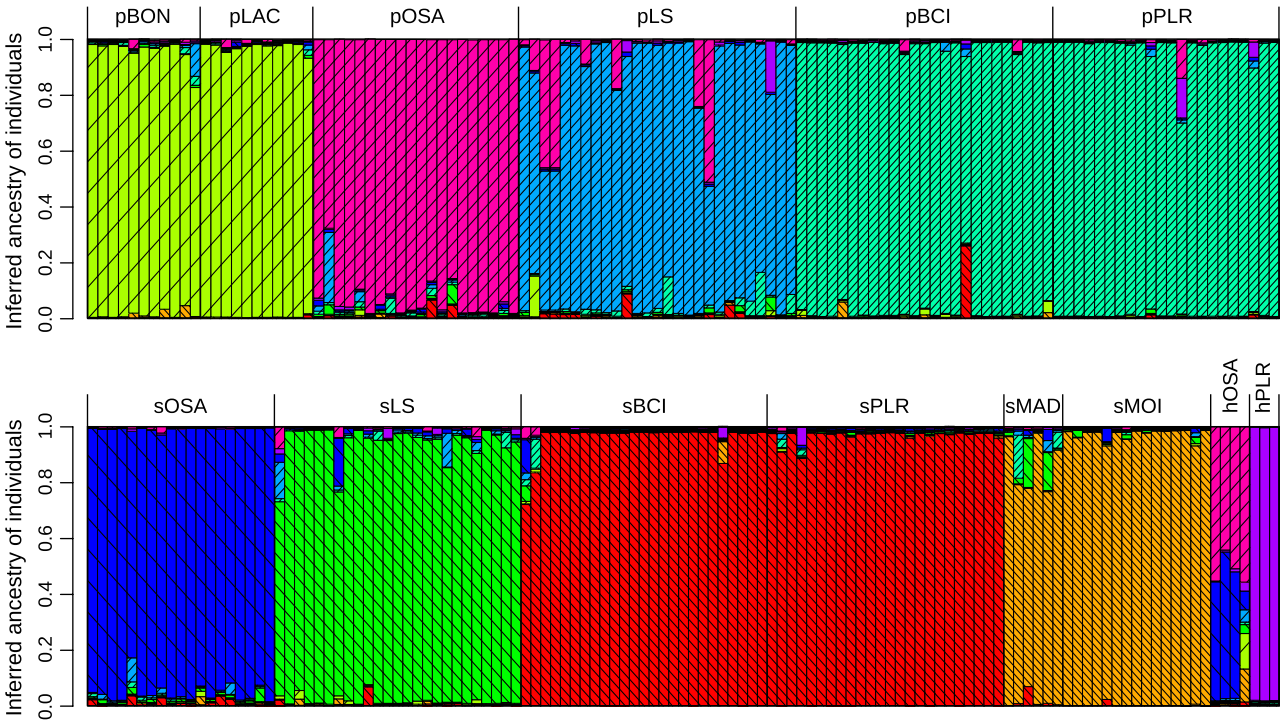

Fig. 2 Genetic structure of *C. pulverulentus* and *C. scaber* estimated by Bayesian assignment with $K = 9$. Each vertical bar represents one individual, and the proportion of each colour within each bar corresponds to the proportion of the individual's genome assigned to each cluster. Individuals are ordered by population, and the codes correspond to the species name (p for *C. pulverulentus*, s for *C. scaber* and h for putative hybrids) followed by the population code (see Fig. 1 and Table 1). The two species are represented on separate lines for clarity, but all the samples have been analysed together.

Our site-by-site STRUCTURE analysis shows low, but detectable, levels of admixture between the species, with 13.9% of individuals across species having 1–5% of their genome assigned to the other species, 2.7% having 5–10% assigned to the other species and 1% having more than 10% assigned to the other species (Tables S3 and S4, Fig. S3, Supporting information). These results are consistent with our overall STRUCTURE analysis (Fig. 2), but are clearer in this site-by-site analysis due to the lower number of clusters involved.

Strong population structure also was observed with the PCA (Fig. 3). The two species were easily distinguished when projecting on the first two principal components. Within species, different patterns were observed between *C. pulverulentus* and *C. scaber*. In the former, two major groups can be distinguished, the first one composed of the individuals from Mexico and the second one by all other individuals. Populations from Central America largely overlapped. *C. scaber* appeared more diverse, with three very distinct groups, formed by individuals from Bolivia, Panama and Costa Rica, respectively. Noticeably, individuals from Panama are very distinct from those from Costa Rica and actually explain the first source of variation (PC1) in the whole data set, indicating very strong differentiation of these populations. Although some overlap among populations is observed on the PCA, the discriminant analyses assigned most individuals to their own population with a very high probability (Fig. 4). The only exceptions were the close-by populations of PLR and BCI in Panama, especially for *C. pulverulentus*.

Discussion

Genetic isolation within species

Our findings of low gene flow among geographically distinct portions of each species range are interesting in the light of the high rate of speciation observed in the genus, especially in topographically complex regions of the Neotropics (Kay *et al.* 2005). For example, the LS and OSA populations, in lowlands on either side of the continental divide at a distance of approximately 220 km, form different STRUCTURE clusters with low admixture. This isolation is similar to what is observed between the Costa Rican and Panamanian populations, with approximately 420 and 480 km distance between BCI and OSA and between BCI and LS, respectively. In contrast, samples from MOI and MAD from the periphery of the Amazon Basin in Bolivia, approximately

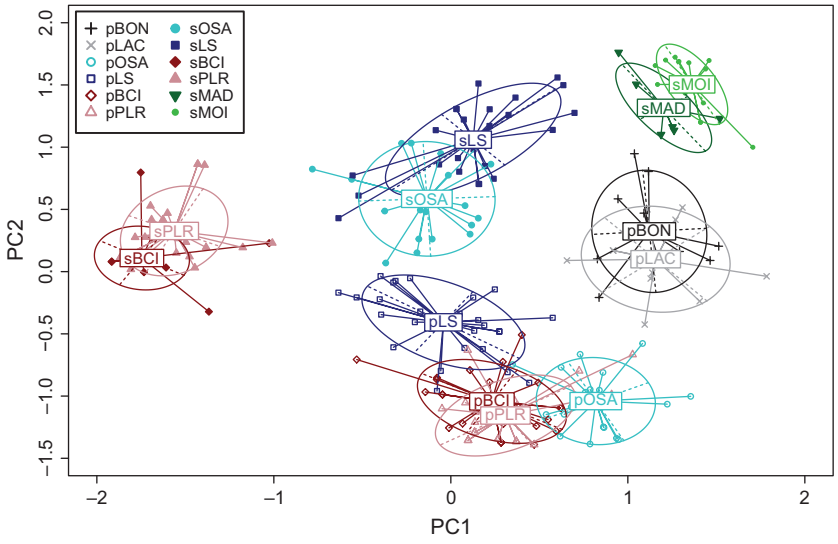

Fig. 3 Genetic structure of *C. pulverulentus* and *C. scaber* estimated by PCA. The first two axes (PC1 & PC2) represent 19.13% and 14.84% of the variation, respectively. Each colour corresponds to a single site, *C. pulverulentus* individuals are represented by open symbols, and *C. scaber* by filled symbols.

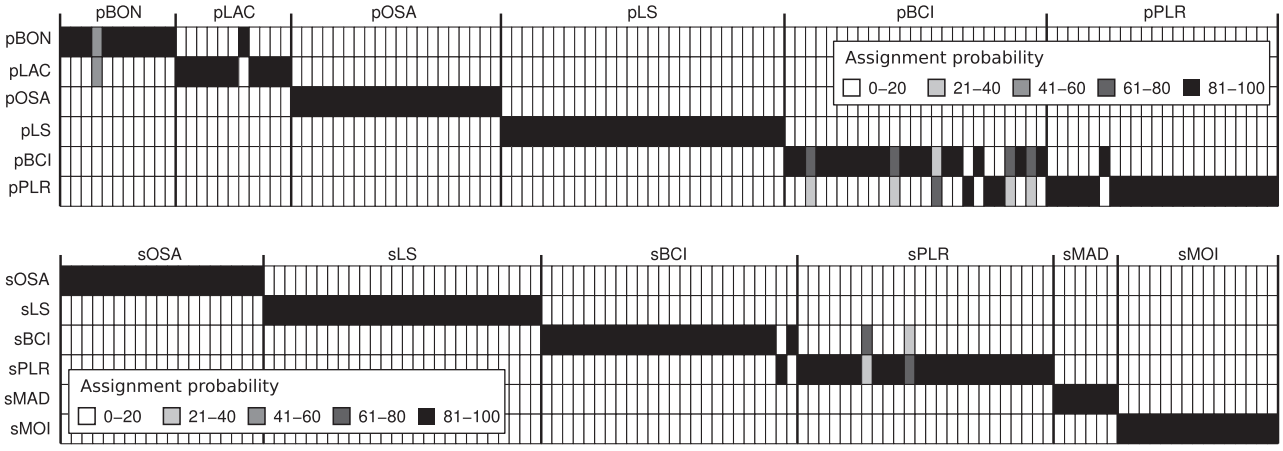

Fig. 4 Genetic structure of *C. pulverulentus* and *C. scaber* estimated by DAPC. Each vertical bar represents one individual (ordered by population with labels as in Fig. 3) and each row a putative population of origin. The shading of each cell corresponds to the assignment probability of a given individual to a given population.

550 km apart, are assigned to the same cluster. Even at very small spatial scales, we are able to distinguish samples from different locations using DAPC. For example, all but two of the samples from LAC and BON, collected approximately 11 km apart, are assigned to the correct population. The only exception is BCI and PLR, also at approximately 11 km distance, for which 5 of 47 total samples were assigned to the other population. At least at putatively neutral loci, it appears that gene flow is restricted enough for populations to diverge along separate evolutionary trajectories. The morphological and ecological similarity of conspecific populations may be maintained by stabilizing selection or through the spread of universally favoured alleles under strong selection (Slatkin 1976; Morjan & Rieseberg 2004). Although many plant species show high F_{ST} and little neutral gene flow among populations (Levin 1979; Morjan & Rieseberg 2004), our study species are predominantly outcrossing and rely on pollinators known to fly long distances on their foraging routes (Stiles 1975, 1978; Stiles & Wolf 1979) and thus may have been expected to show moderate-to-high connectivity among populations. Interestingly, *C. scaber* shows much higher isolation among populations in Costa Rica and Panama than does *C. pulverulentus*, even though they share the same primary pollinator, the Long-billed Hermit hummingbird (*Phaethornis longirostris*), and have no obvious differences in traits related to seed dispersal. This may reflect the fact that *C. scaber* is also partly pollinated by territorial hummingbirds, such as *Amazilia tzacatl* and *Thalurania comlumbica* (Kay & Schemske 2003), that they have contrasting histories of

colonization and range expansion because of other ecological differences or that there is stronger selection against migrants between populations of *C. scaber*.

Although it is becoming increasingly clear that many tropical plant lineages have experienced very rapid and recent bursts of speciation, very little is known about genetic processes of divergence among wild tropical plant populations or gene flow between closely related species, and this study represents one of the few attempts to quantify gene flow at a large geographical scale, especially in herbaceous species. A similar contrast in genetic divergence was found between Mesoamerican and Amazonian populations of the hummingbird-pollinated Neotropical tree *Symphonia globulifera*, with low genetic structure across the Amazon basin and high structure among Mesoamerican and Andean populations (Dick *et al.* 2003). More recently, Jones *et al.* (2012) compared the genetic structure of four different trees across the Isthmus of Panama and identified a sharp contrast between animal-dispersed species displaying a strong genetic structure at a small geographical scale (*S. globulifera* and *Simarouba amara*) and wind-dispersed species (*Jacaranda copaia* and *Luehea seemannii*) with a much lower structure. Although little is known about seed dispersal mechanisms in *Costus*, they exhibit traits consistent with bird dispersal, displaying small white fruits against the bright red interior of the floral bracts. The limited gene flow observed in this study among populations of the two species may have been an important contributing factor to the rapid radiation of the genus *Costus* since its recent colonization of the Neotropics (Kay *et al.* 2005). However, to formally test this hypothesis, it would be necessary to extend this observation to other species of Neotropical *Costus* and compare it to the levels of gene flow within African or Asian taxa that did not go through such a radiation.

Genetic isolation between species

Despite being closely related, occurring sympatrically, and being able to produce hybrids that are vigorous and fertile in the greenhouse (Kay 2006), our data here show that these species are strongly genetically isolated from each other. Excluding apparent morphological hybrids, our geographically structured sampling did not find evidence of the depressed genetic distance between sympatric populations that would be expected under high levels of introgression. Moreover, interspecific admixture of the genomes of sympatric individuals was limited, and all individuals could be confidently assigned to the correct species using DAPC (although not always to the correct population, see above). These results are striking because there is no single isolating mechanism that functions as a complete barrier to hybridization between these species (Kay 2006). The species exhibit subtle differences in habitat affinity, in terms of soil moisture and light availability, but their habitats occur interspersed and their shared pollinator, *P. longirostris*, is known to travel long distances on foraging flights and visit both *Costus* species (Kay 2006). The differences in floral morphology prevent pollen transfer from *C. scaber* to *C. pulverulentus*, but only reduce interspecific transfer in the other direction (Kay 2006). *C. scaber* pistils show partial incompatibility with *C. pulverulentus* pollen, resulting in reduced pollen adhesion, germination and tube growth, but these barriers are all incomplete. Nonetheless, together these multiple incomplete forms of reproductive isolation effectively function as an effective barrier to genetic exchange (Kay 2006).

In line with the low amounts of observed introgression, our testing of putative hybrids shows that a low level of hybridization is occurring, at least in some sympatric regions. Genetic admixture in the four putative hybrids from OSA is consistent with our hypotheses that three represent F1 hybrids between *C. pulverulentus* and *C. scaber* and one represents an F1 hybrid between *C. pulverulentus* and another unsampled sympatric *Costus* species that shares the same pollinator (Kay & Schemske 2003). Although the putative hybrids from PLR formed their own cluster in the STRUCTURE analysis, they were identified as backcrosses with *C. scaber* in the NewHybrids analysis. However, because of the presence of other interfertile *Costus* species at this site, it is not possible to assign these individuals to a particular hybrid category, but only to reject the hypothesis that they are F1 hybrids between our two study species. At 11 of the 15 loci, they had alleles of the two local *C. pulverulentus* and *C. scaber* populations, at 3 additional loci, they had alleles present in other populations of *C. pulverulentus* and *C. scaber*, and at one locus, they had an allele shared with sympatric (and interfertile) *Costus allenii* and another allele that is unique within the Neotropical *Costus* radiation (K. M. Kay & Y. Surget-Groba, unpublished data). Although these plants appeared vigorous, with several inflorescences, we observed that the majority of developing seeds were aborted (K. M. Kay, personal observation).

Taken together, our results suggest that introgression is limited and much of the hybridization may be limited to the formation of F1 hybrids, even though low levels of admixture have been detected in sympatric populations (Tables S2 and S3, Supporting information). Although we have grown hybrids in the greenhouse, postzygotic isolation is not well understood in this system. In the greenhouse, there is some reduction in hybrid seed germination, but once germinated, F1 hybrids are extremely vigorous and exhibit over 90%

pollen fertility as measured by pollen staining (Kay 2006). F2 hybrids and backcrosses are also vigorous in the greenhouse, but rates of pollen fertility vary widely from 3% to 100% fertile grains (K. M. Kay, unpublished data). Hybrid fitness under natural field conditions, however, is largely unknown and may be affected by ecological factors not operating in the greenhouse. For example, hybrid survival may be lower in the field if hybrids are unsuited for the available abiotic habitats (e.g. Hatfield & Schluter 1999) or if hybrids are not well defended against pests and pathogens (Whitham 1989; Fritz *et al.* 1994), which are abundant in the tropical forest. If there is substantial survival selection against hybrids in the field, our sampling of adult plants for this study would lead to underestimation of the rate of hybridization, and it will be interesting to compare our results to levels of hybridization among seeds and seedlings. Hybrid fertility may also be lower in the field if the intermediate floral morphology leads to lower success at exporting or receiving pollen to and from available mates. Based on our understanding of the dynamics of pollen placement on the hummingbirds and pollen transfer between the parental species (Kay 2006), differences in floral morphology between the hybrids and either parent are likely to limit backcrossing to either species, but especially to the longer-flowered *C. pulverulentus*.

A similar study comparing intra- and interspecific gene flow in Neotropical inselberg *Pitcairnia* species found extensive interspecific admixture at sympatric sites and low levels of migration among populations within a species (Palma-Silva *et al.* 2011). In that case, introgression was extensive enough to detect differences between sympatry and allopatry at a very small geographical scale (approximately 20 km). It will be important to replicate these types of studies across many more species at sufficient geographical scales in order to detect commonalities across plant lineages, despite constraints of money, genetic markers and the difficulties of collecting samples across multiple countries with diverse permitting procedures.

Implications for speciation and reinforcement

Our results are important in the context of the pattern of reinforced prezygotic isolation found between sympatric populations of these species. In certain cases of sympatric contact between closely related taxa, hybridization may cause the reinforcement of prezygotic barriers as a late stage in the speciation process (Dobzhansky 1940; Mayr 1959; Grant 1966). Yet, the population genetic conditions under which this can occur are thought to be limited. With extensive gene flow between species, recombination between genes for mating preferences and other species-specific traits should lead to a breakdown in divergence and preclude the evolution of stronger reproductive isolation (Felsenstein 1981; Butlin 1987; Noor 1999; reviewed in Servedio & Noor 2003). In contrast, with only rare hybridization, selection may be too weak to lead to reinforcement (reviewed in Coyne & Orr 2004). Finally, if hybrids are formed but gene flow precluded because of hybrid inviability or sterility, then speciation is already complete and any sympatric divergence is more appropriately considered reproductive character displacement, which can occur between even distantly related species (Butlin 1987). Nevertheless, quantitative empirical studies of interspecific gene flow in empirical systems showing evidence for reinforcement are rare, especially in plants (reviewed in Ortiz-Barrientos *et al.* 2009). In one promising plant system, sympatric flower colour differences reduce hybridization between two *Phlox* species, but hybrids are nearly completely sterile, and the extent of introgression resulting from the limited hybridization is unknown (Hopkins & Rausher 2012; Hopkins 2013).

Between *C. pulverulentus* and *C. scaber*, the floral morphology differences and pollen–pistil incompatibility show geographical patterns consistent with reinforcement (Kay & Schemske 2008), which led us to hypothesize that hybridization and selection have been important in this system. Here, we confirmed genetically that the morphological intermediate individuals found in the field were hybrids (F1 or backcrosses), but that overall hybridization events have led to limited amounts of introgression between the genomes of individuals that do not show morphological evidence of hybridization. When looking at the sympatric populations (Fig. S3, Tables S1 and S2, Supporting information), we observe that several individuals from both species have a small portion of their genome assigned to the other. For instance, across the four sympatric sites, 3.8% of the individuals have more than 5% of their genome assigned to the other species. Thus, our results are consistent with low levels of hybridization and introgression predicted by some reinforcement theory (Kirkpatrick 2000; Kirkpatrick & Ravigne 2002) and provide important empirical data on levels of introgression in reinforced systems.

Given the complex mutation pattern of microsatellites, it is possible that homoplasy could be responsible for part of this pattern. The spatial distribution of alleles can help to distinguish between introgression and homoplasy in some situations. For instance, one allele that is relatively common in the Mexican populations of *C. pulverulentus* is also found in a single individual of *C. scaber* in Bolivia. Because of the distance between these two populations (>3000 km), it is very unlikely that the presence of this allele in C. *scaber* is due to

introgression. It is also unlikely that the presence of this allele in the two species is due to ancestral polymorphism, because it was not found in any other population. In this case, homoplasy is the most likely explanation for the allele sharing. In contrast, several individuals have alleles that are absent from all other individuals of the same species, but common in the other species in sympatry or from close-by populations. For instance, at the LS site, individual one from *C. scaber* (see Fig. S3, Table S2, Supporting information) has four alleles not found in other *C. scaber* individuals but common in the sympatric *C. pulverulentus*, a pattern that can be attributed to introgression. In the future, we plan to extend this study to different markers (SNPs) that have lower substitution rates and are more amenable to coalescent simulations, in order to rigorously test whether there has been a limited amount of interspecific gene flow compatible with the evolution of reproductive isolation by reinforcement.

It is also important to understand the levels of intraspecific gene flow in the context of reinforcement. High levels of intraspecific gene flow from allopatric populations may swamp reinforcement in sympatry (Bigelow 1965), especially if reinforced alleles are disadvantageous elsewhere. Empirical studies have shown that levels of intraspecific gene flow may affect the observed geographical distribution of reinforced alleles but do not prevent reinforcement from occurring (Schaeffer & Miller 1992; Noor *et al.* 2000; Hopkins *et al.* 2011). Here, we show limited levels of intraspecific gene flow between allopatric and sympatric populations of *C. pulverulentus* and *C. scaber* that are consistent with the strong pattern of strengthened prezygotic barriers they exhibit in sympatry. We also show limited intraspecific gene flow among populations within the region of sympatry that are consistent with our site-specific patterns of reinforced prezygotic barriers (Kay & Schemske 2008).

Overall, we find that even with weak intrinsic postzygotic isolation and some hybridization at sympatric sites, introgression appears to be limited. Although we do not know the distribution of these markers across the genome, they are unlikely to be from genes under divergent adaptation and more likely to represent patterns of neutral gene flow. Thus, the low levels of admixture suggest both that hybridization does not frequently proceed beyond the F1 generation, consistent with selection against hybrids, and that intraspecific migration rates among different portions of the geographical ranges of each species are low, consistent with the observed pattern of reinforcement and with further allopatric divergence within species. More explicit modelling of introgression during late stages of speciation will require sequence data with a more comprehensive sampling of the genome. Future work will focus on genomic data and will evaluate genealogies of loci linked and unlinked to QTL underlying different forms of reproductive isolation.

Acknowledgements

This work was facilitated by the Costa Rican Ministerio del Ambiente y Energía, the Panamanian Autoridad Nacional del Ambiente, El Museo de Historia Natural Noel Kempff Mercado in Santa Cruz de la Sierra, Bolivia, and the directors of Amboró and Madidi National Parks in Bolivia. We thank J. Yost, V. Apkenas and M. Bontrager for assistance with laboratory work and four anonymous reviewers for helpful suggestions on improving the manuscript. This work was funded in part by NSF DEB-0947138, a NSF GRF and DDIG, and a Hellman Fellowship to KMK, and YSG was supported by Chinese Academy of Sciences fellowship for young international scientists (grant number 151C53WJQNKXJ20110008). The data analyses were completed with the cooperation of HPC Center, Kunming Institute of Botany, CAS, China.

References

Anderson EC, Thompson EA (2002) A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, **160**, 1217–1229.

Barton N, Bengtsson BO (1986) The barrier to genetic exchange between hybridizing populations. *Heredity*, **57**, 357–376.

Bigelow RS (1965) Hybrid zones and reproductive isolation. *Evolution*, **19**, 449–458.

Butlin R (1987) Speciation by reinforcement. *Trends in Ecology & Evolution*, **2**, 8–13.

Coyne J, Orr HA (2004) *Speciation*. Sinauer, Sunderland, Massachusetts.

Croat TB (1978) *Flora of Barro Colorado Island*. Stanford University Press, Stanford, California.

Dick C, Abdul-Salim K, Bermingham E (2003) Molecular systematic analysis reveals cryptic tertiary diversification of a widespread tropical rain forest tree. *The American Naturalist*, **162**, 691–703.

Dobzhansky T (1940) Speciation as a stage in evolutionary divergence. *The American Naturalist*, **74**, 312–321.

Egan SP, Funk DJ (2009) Ecologically dependent postmating isolation between sympatric host forms of *Neochlamisus bebbianae* leaf beetles. *Proceedings of the National Academy of Sciences*, **106**, 19426–19431.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, **10**, 564–567.

Felsenstein J (1981) Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, **35**, 124–138.

Fritz RS, Nichols-Orians CM, Brunsfeld SJ (1994) Interspecific hybridization of plants and resistance to herbivores: hypotheses, genetics, and variable responses in a diverse herbivore community. *Oecologia*, **97**, 106–117.

Gentry AH (1982) Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Annals of the Missouri Botanical Garden*, **69**, 557–593.

Grant V (1966) The selective origin of incompatibility barriers in the plant genus *Gilia*. *American Naturalist*, **100**, 99–118.

Grant V (1993) Effects of hybridization and selection on floral isolation. *Proceedings of the National Academy of Sciences of the USA*, **90**, 990–993.

Hatfield T, Schluter D (1999) Ecological speciation in sticklebacks: environment-dependent hybrid fitness. *Evolution*, **53**, 866–873.

Hedrick PW (2005) A standardized genetic differentiation measure. *Evolution*, **59**, 1633–1638.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hoorn C, Wesselingh FP, ter Steege H *et al.* (2010) Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, **330**, 927–931.

Hopkins R (2013) Reinforcement in plants. *New Phytologist*, **197**, 1095–1103.

Hopkins R, Rausher MD (2012) Pollinator-mediated selection on flower color allele drives reinforcement. *Science*, **335**, 1090–1092.

Hopkins R, Levin DA, Rausher MD (2011) Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution*, **66**, 469–485.

Ippolito A, Fernandes GW, Holtsford TP (2004) Pollinator preferences for *Nicotiana alata*, *N. forgetiana*, and their F-1 hybrids. *Evolution*, **58**, 2634–2644.

Janzen DH (1967) Why mountain passes are higher in the tropics. *The American Naturalist*, **101**, 233–249.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**, 94.

Jones FA, Ceron-Souza I, Hardesty BT, Dick CW (2012) Genetic evidence of Quaternary demographic changes in four rain forest tree species sampled across the Isthmus of Panama. *Journal of Biogeography*, **40**, 720–731.

Kane NC, King MG, Barker MS *et al.* (2009) Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. *Evolution*, **63**, 2061–2075.

Kay K (2006) Reproductive isolation between two closely related hummingbird-pollinated neotropical gingers. *Evolution*, **60**, 538–552.

Kay KM, Schemske DW (2003) Pollinator assemblages and visitation rates for 11 species of neotropical *Costus* (Costaceae). *Biotropica*, **35**, 198–207.

Kay KM, Schemske DW (2008) Natural selection reinforces speciation in a radiation of Neotropical rainforest plants. *Evolution*, **62**, 2628–2642.

Kay KM, Reeves PA, Olmstead RG, Schemske DW (2005) Rapid speciation and the evolution of hummingbird pollination in neotropical *Costus* subgenus *Costus* (Costaceae): evidence from nrDNA ITS and ETS sequences. *American Journal of Botany*, **92**, 1899–1910.

Kirkpatrick M (2000) Reinforcement and divergence under assortative mating. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **267**, 1649–1655.

Kirkpatrick M, Ravigne V (2002) Speciation by natural and sexual selection: models and experiments. *American Naturalist*, **159**, S22–S35.

Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. *The American Naturalist*, **175**, 316–334.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Levin DA (1979) The nature of plant species. *Science*, **204**, 381–384.

Mayr E (1947) Ecological factors in speciation. *Evolution*, **1**, 263–288.

Mayr E (1959) Isolation as an evolutionary factor. *Proceedings of the American Philosophical Society*, **103**, 221–230.

McBride CS, Singer MC (2010) Field studies reveal strong postmating isolation between ecologically divergent butterfly populations. *PLoS Biology*, **8**, e1000529.

McDade LA, Hartshorn GS (1994) La Selva biological station. In: *La Selva: Ecology and Natural History of a Neotropical Rain Forest* (eds McDade LA, Bawa KS, Hespenheide HA, Hartshorn GS), pp. 6–14. University of Chicago Press, Chicago, Illinois.

Molecular Ecology Resources Primer Development Consortium, Abreu AG, Albaina A *et al.* (2012) Permanent genetic resources added to molecular ecology resources database 1 October 2011-30 November 2011. *Molecular Ecology Resources*, **12**, 374–376.

Morjan CL, Rieseberg LH (2004) How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology*, **13**, 1341–1356.

Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 343–353.

Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.

Noor MAF (1999) Reinforcement and other consequences of sympatry. *Heredity*, **83**, 503–508.

Noor MAF, Schug MD, Aquadro CF (2000) Microsatellite variation in populations of *Drosophila pseudoobscura and Drosophila persimilis*. *Genetical Research*, **75**, 25–35.

Ortiz-Barrientos D, Grealy A, Nosil P (2009) The genetics and ecology of reinforcement: implications for the evolution of prezygotic isolation in sympatry and beyond. *Annals of the New York Academy of Sciences*, **1168**, 156–182.

Palma-Silva C, Wendt T, Pinheiro F *et al.* (2011) Sympatric bromeliad species (*Pitcairnia* spp.) facilitate tests of mechanisms involved in species cohesion and reproductive isolation in Neotropical inselbergs. *Molecular Ecology*, **20**, 3185–3201.

Pritchard J, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

R Core Development Team (2011) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey J, Bradshaw HD, Schemske DW (2003) Components of reproductive isolation between the monkeyflowers *Mimulus lewisii* and *M. cardinalis* (Phrymaceae). *Evolution*, **57**, 1520–1534.

Renaut S, Maillet N, Normandeau E *et al.* (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 354–363.

Rousset F (1997) Genetic differentiation and estimation of gene flow from F-Statistics under isolation by distance. *Genetics*, **145**, 1219–1228.

Rundle HD, Whitlock MC (2001) A genetic interpretation of ecologically dependent isolation. *Evolution*, **55**, 198–201.

Schaeffer SW, Miller EL (1992) Estimates of gene flow in *Drosophila pseudoobscura* determined from nucleotide sequence analysis of the alcohol dehydrogenase region. *Genetics*, **132**, 471–480.

Schemske DW (2002) Ecological and evolutionary perspectives on the origins of tropical diversity. In: *Foundations of Tropical Forest Biology: Classic Papers with Commentaries* (eds Chazdon RL, Whitmore TC), pp. 163–173. University of Chicago Press, Chicago.

Schemske DW (2010) Adaptation and the origin of species. *American Naturalist*, **176**, S4–S25.

Seehausen O, Terai Y, Magalhaes IS *et al.* (2008) Speciation through sensory drive in cichlid fish. *Nature*, **455**, 620–626.

Servedio MR, Noor MA (2003) The role of reinforcement in speciation: theory and data. *Annual Review of Ecology Evolution and Systematics*, **34**, 339–364.

Slatkin M (1976) The rate of spread of an advantageous allele in a subdivided population. In: *Population Genetics and Ecology* (eds Karlin S, Nevo E), pp. 767–780. Academic Press, New York.

Smith SD, Baum DA (2006) Phylogenetics of the florally diverse Andean clade Iochrominae (Solanaceae). *American Journal Of Botany*, **93**, 1140–1153.

Stiles FG (1975) Ecology, flowering phenology, and hummingbird pollination of some Costa Rican *Heliconia* species. *Ecology*, **56**, 285–301.

Stiles FG (1978) Ecological and evolutionary implications of bird pollination. *American Zoologist*, **18**, 715–727.

Stiles FG, Wolf LL (1979) *Ecology and Evolution of Lek Mating Behavior in the Long-Tailed Hermit Hummingbird*. The American Ornithologists' Union, Washington, District of Columbia.

Thioulouse J, Chessel D, Doledec S, Olivier JM (1997) ADE-4: a multivariate analysis and graphical display software. *Statistics and Computing*, **7**, 75–83.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic Islands of Speciation in *Anopheles gambiae*. *PLoS Biology*, **3**, e285.

Venables WN, Ripley BD (2002) *Venables: MASS: Modern Applied Statistics With S*.

Whitham TG (1989) Plant hybrid zones as sinks for pests. *Science*, **244**, 1490–1493.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

Y.S.G. and K.M.K. designed and performed research, analysed data and wrote the article. K.M.K. made all field collections.

Data accessibility

Sample locations, Structure and NewHybrids input files and microsatellite data: DRYAD entry doi:10.5061/dryad.8q34c.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Pairwise geographic distances (in km) between the populations studied.

Table S2 Pairwise F_{ST} (above diagonal), and G'_{ST} (below diagonal).

Table S3 Number (and percentage) of individuals from each species having a given percentage of their genome assigned to the other species.

Table S4 Inferred ancestry to the other species for the admixed individuals.

Fig. S1 Isolation by distance plots within and between species showing the relationship between genetic [linearized G'_{ST}: $G'_{ST}/(1-G'_{ST})$] and geographic (log km) distances.

Fig. S2 STRUCTURE results: estimated probability of data L(K) (black squares), and ΔK (black diamonds) for a number of cluster K ranging from 1 to 16.

Fig. S3 Genetic structure of *C. pulverulentus* and *C. scaber* sympatric populations estimated by Bayesian assignment with $K = 2$.