OPEN

A Biogeographic Barrier Test Reveals a Strong Genetic Structure for a Canopy-Emergent Amazon Tree Species

Alison G. Nazareno[1]*, Christopher W. Dick[2] & Lúcia G. Lohmann[1]*

Wallace's (1854) Riverine Barrier hypothesis is one of the earliest explanations for Amazon biotic diversification. Despite the importance of this hypothesis for explaining speciation in some animal groups, it has not been studied extensively for plant species. In this study we use a prominent Amazon tree, *Buchenavia oxycarpa* (Mart.) Eichler (Combretaceae), to evaluate Wallace's hypothesis along the Rio Negro, a major Amazon tributary that has driven allopatric speciation for several animal taxa. We sampled six individuals from sixteen localities along both river banks, and used a modified ddRADseq protocol to identify SNP markers. Our population genomic data revealed strong genetic structure for *B. oxycarpa* sampled across banks of the Rio Negro ($\phi_{CT} = 0.576$, $P < 0.001$), supporting the hypothesis that the Rio Negro acted as a significant genetic barrier for *B. oxycarpa*. Our study shows that gene flow for this large and well-dispersed Amazon tree is impeded by riverine barriers, though this has not yet resulted in speciation. Future studies focused on species with different life histories, including species restricted to non-flooded forests, are needed to further advance our understanding of Amazon rivers as drivers of biotic diversification.

The patterns of population structure can provide insights into the evolutionary history of individual taxa. Migration, a key evolutionary force, is expected to weaken the degree of genetic structure within a species range[1–3], and is largely dependent on species-specific ecological traits[3–6]. For instance, ecological differences among species explains much of the interspecific variance in levels of genetic differentiation in birds across physical barriers[6]. These findings were also observed for other animal species[7–10].

For sessile organisms such as plants, landscape features and life-history traits (e.g., associated with pollination and dispersal) can restrict migration and increase the amount of genetic variation among populations[3,4,11,12]. Indeed, strong genetic differentiation has been observed across physical barriers for plant species with restricted dispersal abilities[13–15]. For example, highland and mountain ranges acted as effective barriers to gene flow for the tree species *Styrax sumatrana*[14], a species rarely dispersed by animals[16]. On the other hand, for plant species with high dispersal abilities, geographic barriers may not promote significant population structure[17,18]. In the Amazon Basin, where rivers are a potential cause of allopatric population differentiation for a plethora of taxa, high rates of cross-river gene flow have been documented for several different plant species[17–19]. Along the Rio Negro, one of the largest tributaries of the Amazon river broadly acknowledged as a biogeographic barrier since Wallace[20], the first study to test this hypothesis for plants reported high rates of gene flow across river banks for two animal dispersed canopy-emergent trees[17]. A more recent population genomics study recovered weak genetic differentiation of a fish-dispersed shrub species spanning the Rio Negro[18]. Thus, extensive gene flow across large Amazon rivers may be expected for plant species with high dispersal abilities[17–19,21]. One limitation of most prior studies is that they involved species that are most common in the flooded forests and therefore able to endure flooding and benefit from seed dispersal by water. There are as yet very few studies of species that occur primarily in *terra firme* (unflooded) forest[17].

The study species *Buchenavia oxycarpa* (Mart.) Eichler is a prominent timber tree species (up to 30 m in height) distributed in lowland *terra firme* forests of South America (Ecuador, Peru, Brazil, Venezuela and Bolivia).

[1]Departamento de Botânica, Universidade de São Paulo, São Paulo, SP, Brazil. [2]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. *email: alison_nazareno@yahoo.com.br; llohmann@usp.br

1

In Brazil, *B. oxycarpa* is found in the Amazon, Cerrado, and the Atlantic Forest[22]. In the Amazon Basin, *B. oxycarpa* occurs at relatively high densities in humid forests, often near rivers and periodically flooded terrains. Due to its high wood quality, *B. oxycarpa* has been overharvested in some areas, as in the Cerrado[23]. *Buchenavia oxycarpa* is pollinated by small bees[24]. In the Peruvian Amazon, *B. oxycarpa* is dispersed by two sympatric primate species, *Saguinus mystax* and *S. fuscicollis*[25]. Although these primate species are not found in the Central Amazon basin, it is very likely that seeds of *B. oxycarpa* are dispersed by other species of the genus *Saimiri*. The spongy mesocarp of the indehiscent fruits of *B. oxycarpa* may be associated with hydrochory, so water dispersal may be possible for this species[26].

Even though no genetic information is available for *B. oxycarpa*, we would expect high rates of gene flow between populations of this animal-dispersed plant species based on genetic studies of tree species with similar ecology[17–19,21]. As a consequence, populations of *B. oxycarpa* located on opposite margins of the Rio Negro – a putative riverine barrier – should present weak population genetic structure. This picture may be expected since results from recent studies have indicated that long-dispersal syndromes, such as vertebrate-dispersal, are responsible for low levels of genetic differentiation for riverine plant species with widespread distributions[17].

In this study, we aim to test the influence of the Rio Negro as a biogeographic barrier to gene flow for *B. oxycarpa* (i.e., a test of the Wallace's Riverine Barrier hypothesis[20]). While most Amazon tree populations occur in low densities or have scattered distributions, *B. oxycarpa* is among the few plant species that are sufficiently common along both banks of the Rio Negro, allowing an adequate sampling to test whether the Rio Negro is an effective genetic barrier. To this end, we sampled *B. oxycarpa* at sixteen localities on opposite river-banks of the Rio Negro and used a modified high-throughput DNA sequencing methodology (i.e., double-digest Restriction-site associated DNA sequencing – ddRADseq)[27] to identify anonymous SNP markers from the nuclear genome. Through this study, we hope to contribute to our current understanding about the importance of Amazon rivers for the assembly and evolution of the plant biota.

Results

The number of single-end raw reads of 101 bp produced for each lane of HiSeq. 2000 Illumina containing 48 *B. oxycarpa* individuals ranged from 102 million to 154 million. Each read started with a barcode sequence that identified each sample (up to 10 bp long), which was followed by a 6 bp restriction site, and 85 bp of usable data. The mean number of retained reads that passed the default quality filters, including a Phred quality score > 33, and which contained an identifiable barcode, were $2,165,480 \pm 124,807$ SE. Throughout the *B. oxycarpa* genomes, further filtering (10-fold coverage; presence in at least 85% of the individuals; MAF > 0.01) resulted in 3,298 unlinked polymorphic SNP markers with a mean coverage depth per locus of 18.2 ± 5.6x. The minor allele frequency (MAF) averaged 0.095 ± 0.0363 SD. No significant departures from HWE were observed in any population or locus after a Bonferroni adjustment ($P > 0.000015$). In addition, no LD was observed after a sequential Bonferroni correction for k tests ($k = 5.4 \times 10^6$, $P < 9.2 \times 10^{-9}$).

We detected 278 potential loci that were under diversifying selection with the false discovery rate (FDR) set to 0.05. Thus, a total of 3,020 neutral loci were used in genomic analyses.

Population genomic structure and the genetic barrier hypothesis.

Using the MDS and Bayesian clustering methods, we identified one potential barrier to gene flow in the Rio Negro for both sets of SNP loci analyzed (i.e., neutral loci and loci putatively under divergent selection). An examination of the stress values (Kruskal's stress $= 0.003$ for neutral loci and 0.010 for loci putatively under divergent selection) determined that two dimensions were sufficient to explain genetic patterns along the Rio Negro. The MDS plot indicated a clear separation of samples between the left and right banks of the Rio Negro (Fig. 1A,B). The genetic structure pattern from the MDS analysis closely matched that obtained using Bayesian clustering analysis. Geneland results clearly delineated two groups with minimal variance in the posterior probabilities of population estimation over multiple runs using a spatial model (Fig. 2A,B).

The 120 pairwise estimates of F_{ST} based on neutral loci ranged from 0.0086 to 0.6711 and all but 17 pairwise estimates were statistically significant ($P < 0.05$), indicating limited differentiation between *B. oxycarpa* population pairs from the Rio Negro (see Table S2). Based on the results of the multiple matrix regression using neutral loci exclusively, a significant portion of the variation in pairwise genetic distances is explained mainly by the historic isolation caused by the river acting as a barrier to gene flow (93.2%). Part of the genetic divergence remains unexplained (6.7%), and only a small proportion (0.1%) of that variation can be explained by the combined effect of a river barrier plus isolation by distance. Results of simple matrix correlation between genetic and geographic distance were not significant when applied separately to both banks of the Rio Negro (r $= 0.274$, $P = 0.191$ for the right bank; r $= 0.067$, $P = 0.740$ for the left bank) nor when applied between pairs of *B. oxycarpa* populations on opposite banks of the river (r $= 0.055$, $P = 0.663$).

The hierarchical multi-locus evaluation of genetic differentiation, performed using an AMOVA, agrees with our previous results indicating that the Rio Negro can be a barrier to gene flow. Most of the genetic variation was attributable to differences observed between banks ($\phi_{CT} = 0.576$, $P < 0.001$; Table 1) rather than among populations within river banks ($\phi_{SC} = 0.027$, $P < 0.001$, Table 1), strengthening our findings that Rio Negro is a genetic barrier for *B. oxycarpa*.

Population tree and historical migration events.

The best maximum likelihood tree showed that populations on each side of the Rio Negro form two clades that correspond to the left and right river bank samples (Fig. 3A), explaining 99.49% (Fig. 3A) of the variance in relatedness among *B. oxycarpa* populations in the Rio Negro. Historical migration events (up to three events) were added sequentially to the tree. The graph model (Fig. 3B) explained 99.61% of the variance in relatedness among *B. oxycarpa* populations. The first added migration
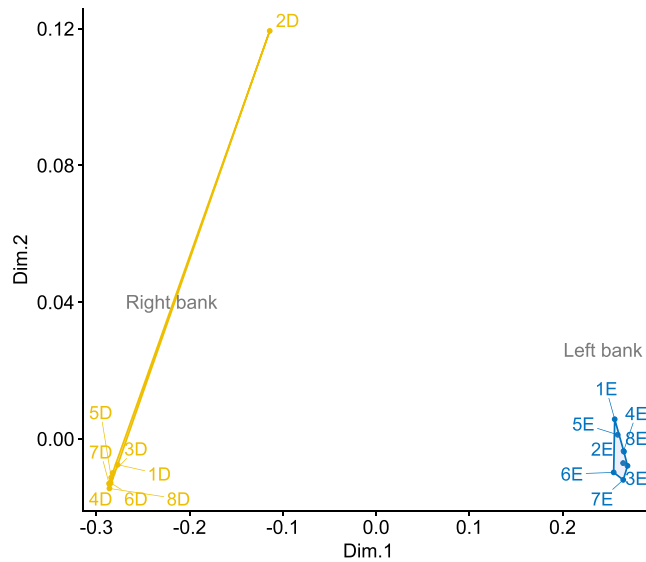
Figure 1. Patterns of geographical structure in *Buchenavia oxycarpa* (Mart.) Eichler represented by multidimensional scaling (MDS) of the matrix of genetic distances based on 3,020 neutral loci.
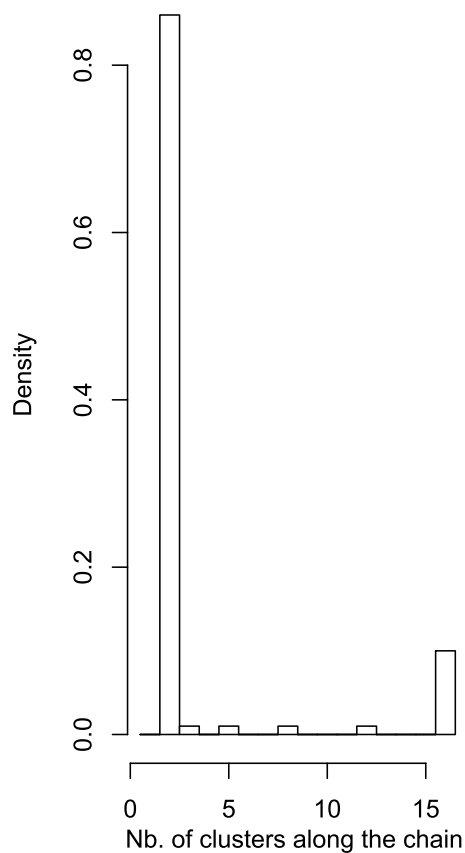

Figure 2. Population clustering analyses for *Buchenavia oxycarpa* (Mart.) Eichler as calculated by GENELAND showing the density of the estimate of k along the Markov chain (after a burn-in of 1000 × 100 iterations), using 3,020 neutral loci.

edge goes from Pop7L to Pop6L with a weight of 0.43 (Fig. 3B). Although there are migrations between *B. oxycarpa* populations within and among banks of the Rio Negro (Fig. 3B), migration events with high weights were observed for population pairs on the same bank of the Rio Negro (e.g., Pop3L-Pop7L and Pop7L-Pop6L; Fig. 3B).

	Sum of squares	Variance components	% of Variation	P-value
Between banks	14018.14	181.86	57.67	0.000
Among populations within banks	2858.94	8.48	2.69	0.000
Within populations	16881.43	125.01	39.64	0.000
Total	33758.51	315.35		

Table 1. Analysis of Molecular Variance (AMOVA) based on 3,020 neutral loci for *Buchenavia oxycarpa* (Mart.) Eichler along the Rio Negro, Amazon Basin, Brazil.
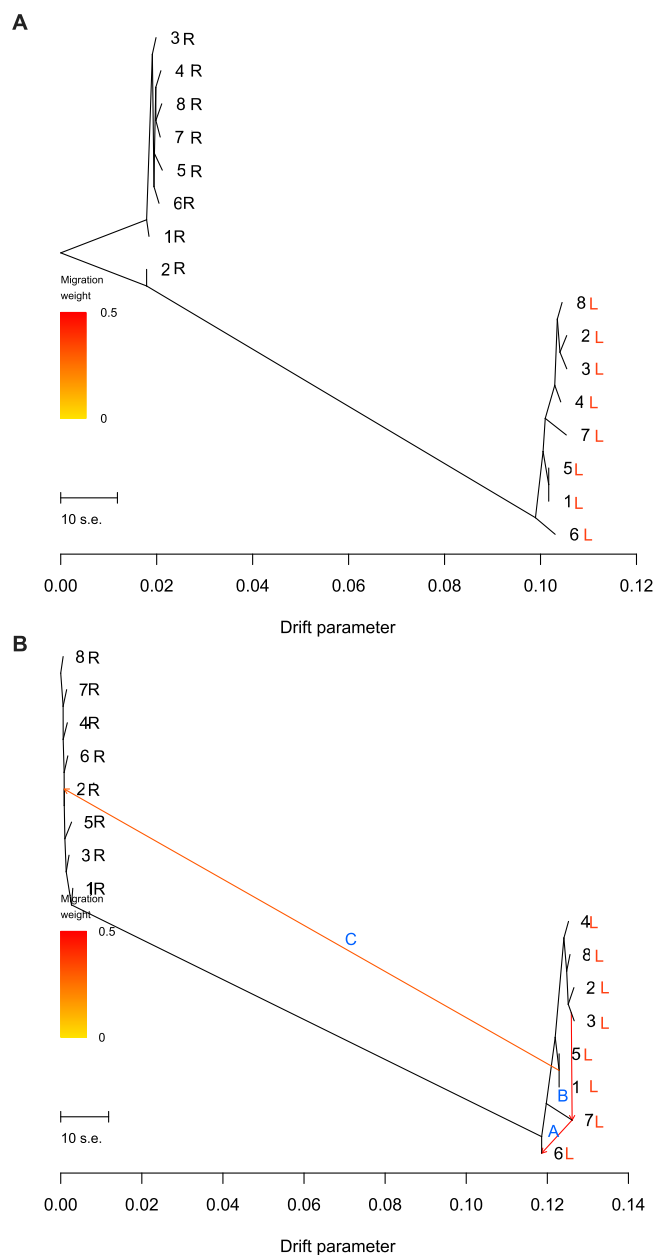
Figure 3. (**A**) Structure of the graph inferred by TreeMix for *Buchenavia oxycarpa* (Mart.) Eichler populations along the left (L) and right (R) banks of the Rio Negro, Amazon Basin, Brazil. (**B**) Population tree with branch lengths scaled to the amount of genetic drift between banks of the Rio Negro. Arrows indicate the inferred migration proportion and the letters, in blue, indicate the order of the migrations. The scale bar shows ten times the average standard error of the entries in the sample covariance matrix. The drift parameter reflects the amount of genetic drift that has occurred among *B. oxycarpa* populations.

Discussion

Wallace's Riverine Barrier hypothesis[20] is an allopatric speciation model that aims to explain contemporary biodiversity patterns in the Amazon basin. From a reinterpretation of the Wallace's Riverine Barrier hypothesis in population genetics terms, we can expect that large rivers, such as the Amazon, will reduce or prevent gene flow between populations on opposite river banks, ultimately leading to allopatry and species distributions that are restricted to particular interfluvial regions. A large body of studies have shown that the rivers of the Amazon Basin are important barriers to gene flow for birds[28,29], amphibians[30], primates[31], and plants[18]. Although biogeographic hypotheses such as Wallace's have often been evoked to explain diversity patterns, plants do not show the same patterns of endemism as animals[32]– leaving a major gap in our understanding of the evolutionary history of Amazon forests[33]. Furthermore, basic knowledge of the Amazon landscape remains elusive[34]. Therefore, an improved understanding of the genetic structure of Amazonian plants can help advance our understanding of the region's geophysical history.

In this study, we tested whether the Rio Negro, a major tributary of the Amazon river and an allopatric barrier for primates, birds and amphibians[28,30,31], is a barrier to gene flow for the tree species *Buchenavia oxycarpa*. While most Amazon tree populations occur in low densities or have scattered distributions[35], *B. oxycarpa* is among the few plant species that are sufficiently common to allow adequate sampling along both banks of the Rio Negro. We used RadSeq to identify neutral SNP loci, which allowed for robust genetic analyses and a characterization of population genetic structure despite the relatively small population sample sizes[27].

Contrary to what may be expected for an animal-dispersed plant species[4], our population genetic data indicated a strong and significant population genetic structure for *B. oxycarpa* across the Rio Negro. The Bayesian and genetic distance–based clustering analyses grouped the populations located at opposite banks of the Rio Negro into two different groups (Figs. 1 and 2), showing a spatial pattern that corresponds to geographic location. All populations located on the left bank of the Rio Negro were grouped within the same cluster, while all populations located on the right bank were grouped within another cluster. These results indicate that historical gene flow via seeds and/or pollen occurred in very low frequency across banks of the Rio Negro. The AMOVA analyses reiterated these results, as the high proportion of total neutral variance attributed to the variance among banks (57.67%) indicated limited gene flow between populations on opposite river-banks separated by up to 11.7 km. These results strengthen the role of the Rio Negro as a historical barrier for *B. oxycarpa*.

The multiple matrix regression analysis using neutral SNPs indicates that 93.2% of the variation in pairwise genetic distances is explained by historical isolation caused by the riverine barrier. This analysis also indicates that geographic distance is not a primary driver of population structure for *B. oxycarpa*. In addition, no similarities between neutral genetic and geographic distance were observed within each bank of the Rio Negro, indicating that divergence with significant gene flow is feasible within each river-bank. In fact, pollen and seeds of *B. oxycarpa* seem to have been dispersed longer distances within than across river banks, with historical gene flow occurring up to 84 km at the left bank of the Rio Negro as indicated by our population graph analysis. This genetic pattern (i.e., no isolation by distance and extensive gene flow within each river bank) was also observed for *Amphirrhox longifolia* in a recent study conducted in the same location[18].

Riverine plant species with low dispersal abilities are more likely to show hierarchical genetic structure[36,37] than plants that can move extensively across rivers[17,18]. The patterns of neutral genetic structure observed here seem to have resulted from species-specific traits. However, this trend is not consistent with what has been observed in other animal-dispersed plant species in wider Amazon rivers. As a matter of fact, the Rio Negro does not seem to pose a barrier for a low-density and widely distributed canopy-emergent tree species (*Caryocar villosum*) that grows in *tierra firme* forests, nor to a habitat-specific tree (*C. microcarpum*) that grows in seasonally flooded black-water forests[17]. These results are expected given the long distances of gene flow associated with bat-pollination and seed dispersal by fish[17].

In this context, the behavior and movement of frugivorous animals seems to constitute an important evolutionary force shaping patterns of genetic structure of riverine plant species. Although the specific disperser of *B. oxycarpa* is unknown in the study area, their seeds are dispersed by two sympatric primate species (*Saguinus mystax* and *S. fuscicollis*) in the Peruvian Amazon[25]. These findings suggest that other primate species (e.g., *Saimiri* spp.) may act as dispersal vectors of the seeds of *B. oxycarpa* along the Rio Negro. Although some primates can eventually cross smaller Amazon rivers (e.g., *Alouatta macconnelli*, *Sapajus apella*[31]), the Rio Negro seems to be a dispersal barrier for primates[20,31]. Indeed, Boubli and co-authors[31] showed that this wider river represents a genetic barrier that limits the distribution of the primate genera *Ateles* and *Chiropotes* to the left bank and *Saguinus* and *Sapajus* to the right bank in the upper Rio Negro. Although water can disperse seeds to greater distances than other dispersal mechanisms[38], the strong genetic differentiation observed for *B. oxycarpa* in opposite banks of the Rio Negro are congruent with the distribution patterns of primates. Further studies of the seed ecology of *B. oxycarpa* may confirm this observation. Studies of this nature would also shed light on the role of plant-animal dispersal mutualisms in shaping patterns of population genetic structure. Given that fruits are a key component of primate diet in the Amazon Basin[39] and that primates are important seed dispersal vectors in tropical forests[40,41], an improved understanding of the actual role of rivers as barriers in this region will further improve our understanding of the role of allopatric speciation for the origin and maintenance of Amazon plant diversity.

Landscape features may also play an important role in structuring genetic diversity of *B. oxycarpa*. The Rio Negro is subject to long periods of flooding[42], especially in the lower portions where *B. oxycarpa* populations were sampled. In addition, inundation extent can vary among river banks, which can lead to differences in flowering and fruiting phenology, possibly restricting gene flow by pollen across river banks. Future studies evaluating the reproductive biology of *B. oxycarpa* in both banks of the Rio Negro may enable us to better predict genetic patterns for this species. Furthermore, river characteristics such the degree of river discharge and width can impact gene flow[18,43,44]. For instance, a recent population genomic study carried out in the core Amazon Basin showed that the strength of the riverine barrier for the plant species *Amphirrhox longifolia* depended on the width of the rivers separating populations[18].

Although the environmental conditions, morphological, and reproductive traits of *B. oxycarpa* individuals were not recorded in our study, these types of data can help infer explicit targets of selection and adaptation, especially when correlated with allelic frequencies. Studies of this nature would be extremely valuable for species of conservation concern such as *B. oxycarpa*, which has been drastically overharvested in some regions[23].

Overall, the population genetic structure data reported here supported the hypothesis that the Rio Negro acted as a significant genetic barrier for *B. oxycarpa*, though not enough to generate novel taxonomic diversity. Although this study provided new evidence about the role of rivers in shaping genetic structure on plants, our knowledge about landscape genetics and allopatric speciation history of Amazon plant diversity is still limited. Clearly, larger-scale population genomic studies on plants with varied ecological traits are greatly needed in order to test whether the patterns observed here are representative of the floodplain flora of the Rio Negro as a whole. Additional studies, conducted in Amazonian waterways with distinct features would complement our current understanding, which is focused on patterns observed along the Rio Negro. It would also be important to study plants that are restricted to the non-flooded *terra firme*, without any adaptations to flood tolerance. While this study focused on the analysis of historical genetic structure, detailed analyses based on contemporary gene flow would provide additional insights. More comprehensive studies are also needed to identify putative associations between ecological-environmental traits and allelic frequency. Studies of this nature would allow us to better understand how deterministic and neutral processes have contributed for the evolutionary trajectory of Amazon plant species.

Materials and Methods

Study area and sampling. The Rio Negro (Fig. 4) is the fifth-largest river in the world[45]. This river runs for 1700 km throughout the Amazon Basin, with the mouth of the river near the city of Manaus, Amazonas State, Brazil. Although the geological history of this river is poorly understood, geomorphological studies suggest that the mouth of the river migrated 150 km eastward to its current location[46]. The floodplain forests bordering the river, known as *igapó*, are home to several hundred tree species[47,48] and are mainly lowland rainforests. Sub-montane, montane, and other lowland vegetation, such as white sand-soil forests[49], also occur throughout the basin. The region has a mean annual rainfall of 2200 mm (IBGE) and the climate is classified as tropical equatorial, with a dry season occurring between June and August, and the rest of the year considered the rainy season. Because its nutrient-poor soils are unsuitable for agriculture, deforestation along the Rio Negro has been minimal.

Leaf samples of flowering *B. oxycarpa* individuals were collected during the rainy season of 2016. Sixteen sampling sites (populations) located on both the left (L) and right (R) banks of the Rio Negro were included in the study (Table S1, Fig. 4). All samples were taken from floodplains, with populations occurring within a 141 km range. Each population was paired with a corresponding population on the opposite river bank, located at distances ranging from 3.3 km (Pop3R – Pop3L, Fig. 4) to 11.7 km (Pop1R – Pop1L, Fig. 4). In each population, six individuals were sampled at intervals of at least 50 m to avoid collection from related trees. One voucher specimen was collected per population (Table S1), and all vouchers were deposited in the University of São Paulo Herbarium (SPF), São Paulo, Brazil. Because this study uses SNPs that offer high-resolution assessment of genetic structure[27,50–53], sample sizes smaller than those normally used in population genetics studies are sufficient as the limitations inherent to small samples are offset by the large numbers of SNPs[18,54]. To validate the use of a small sample size for *B. oxycarpa*, following Nazareno et al.[18], we assessed pairwise genetic differentiation by randomly reducing the number of samples from six to two. The results confirm that the sample sizes used are adequate for genetic structure analyses [F_{ST} (n = 6) = 0.343, 95% CI (0.292, 0.395); F_{ST} (n = 2) = 0.321, 95% CI (0.267, 0.376)].

Library preparation and sequencing. The preparation of the libraries and DNA sequencing followed the protocol described in Nazareno et al.[18]. Briefly, we used the Macherey-Nagel kit (Macherey-Nagel GmbH & Co. KG) to extract genomic DNA from leaf samples. Two genomic libraries were constructed using a double digest RADseq protocol[55] and following the adjustments outlined in Nazareno et al.[18] related to minimizing variation in number of reads per individual. Before digestion, we used the Qubit dsDNA Assay Kit (Invitrogen) to quantify the concentrations of double-stranded DNA. Samples were adjusted to equal molar concentration, and the final DNA concentration for each sample was 500 ng.μL[1]. Two restriction enzymes, *EcoRI* and *MseI* (New England Biolabs), were used to digest each sample, and digestions were carried out in a total volume of 20 μL, using 17 μL of resuspended DNA, 5 units of EcoRI, 5 units of MseI, and 1X CutSmart buffer (New England Biolabs). The protocol consisted of 3 hours at 37 °C, with a final 20 min deactivation step at 65 °C. The Agencourt AMPure XP system (Beckman Coulter) was used to purify the reactions, following the manufacturer's instructions, with elution in 40 μL TE buffer. The cleaned digests were quantified using Qubit to standardize the initial DNA mass to be added into an adapter ligation. Adapter ligations were conducted using a total volume of 30 μL, with 42 ng DNA, 0.22 μM of a non-sample specific *MseI* adaptor (common for all samples), 0.33 μM of a sample specific *EcoRI* double-strand adaptor for each DNA sample, 1U of T4 DNA ligase (New England BioLabs), and 1.3 X T4 ligase buffer which were incubated at 23 °C for 30 min. Restriction enzymes were then heat-killed at 65 °C for 10 min followed by a slow cooling to room temperature (23 °C).

After cleaning the reactions, ligation products were amplified in 20 μL PCRs, containing 13.5 μL of the ligation product, 0.2 μM of Illumina PCR primers, 0.2 mM dNTPs, 1.0 mM MgCl$_2$, 0.5 U of iProof™ High-Fidelity DNA polymerase (BIO-RAD), and 2X of iProof buffer. An Eppendorf PCR System was used for PCR using the following protocol: 98 °C for 30 s, 20 cycles of 98 °C for 20 s, 60 °C for 30 s, and 72 °C for 40 s, followed by a final extension at 72 °C for 10 min. Before pooling samples in each library, samples were purified using the Agencourt AMPure XP system and the DNA quantified using Qubit. The concentration of DNA for each sample ranged from 2.13 ng.μL^{-1} to 13.00 ng.μL^{-1}. We prepared multiplexed libraries with generally the same amount of DNA per sample. The target size range to select genomic fragments was 375–475 bp, and automated size-selection was
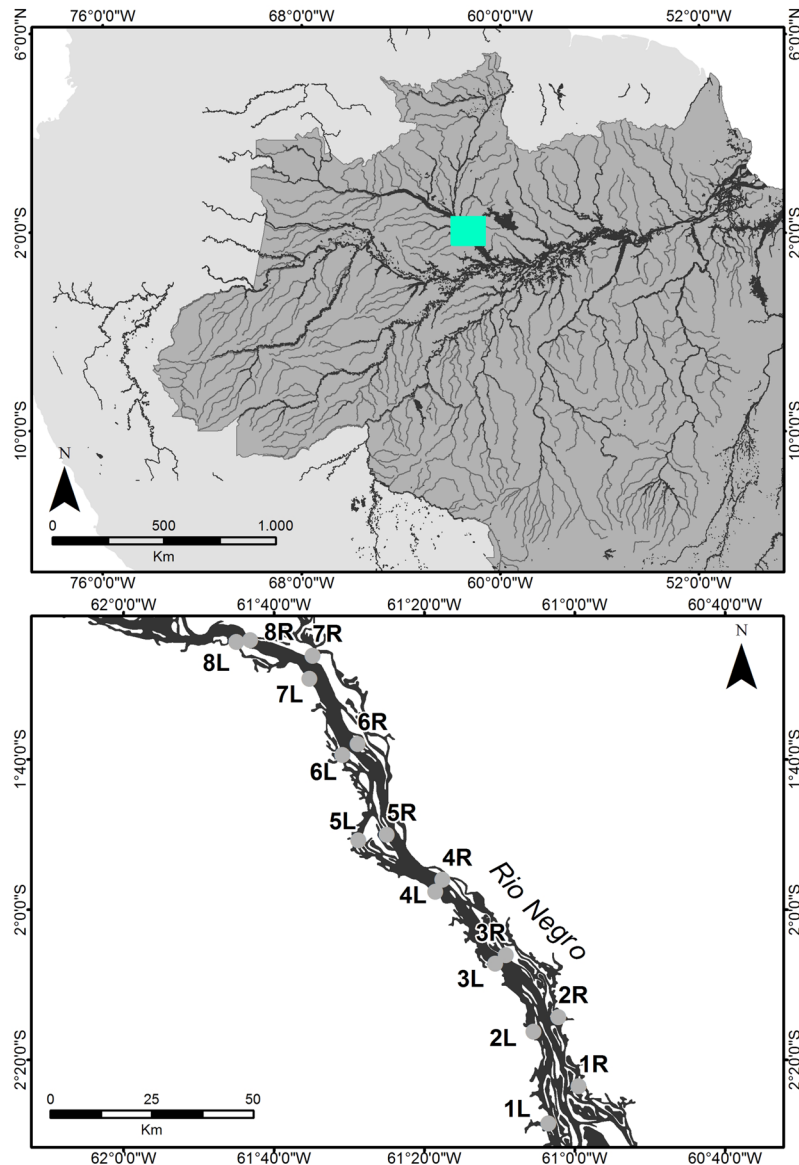
Figure 4. Sampling locations (1 R–8 R, 1 L–8 L) of *Buchenavia oxycarpa* (Mart.) Eichler along the left (west) and right (east) banks of the Rio Negro, Amazon Basin, Brazil.

performed using a 2% agarose cartridge (Pippin Prep; Sage Science, Beverly, MA). We used the Agilent 2100 Bioanalyzer (Agilent Technologies) with the Agilent DNA 1000 Kit to measure size, quantity and quality of each individual library. Libraries were sequenced (100-bp single-end reads) at The Centre for Applied Genomics in Toronto, Canada, in a single lane (each pooled with 48 *B. oxycarpa* individuals) of an Illumina HiSeq. 2000 flow cell (Illumina Inc., San Diego, CA).

Identifying and genotyping SNPs. The procedure used to identify and genotype SNPs is fully outlined in Nazareno *et al*.[18]. The program Stacks 1.35[56,57] with the *de novo* assembly was used to analyze files containing the raw sequence reads. The program ustacks was used to produce consensus sequences of RAD tags, as it aligns short-read sequences from a single sample into exactly matching stacks. To estimate the diploid genotype for each individual at each nucleotide position, we used a maximum-likelihood framework[58]. The optimal minimum depth of coverage to create a stack was set to three sequences, with the maximum distance permitted between stacks set to two nucleotides, and the maximum number of stacks allowed per *de novo* locus set to three. We used an alpha value for the SNP model of 0.05. A catalog of consensus loci containing all loci and merging all alleles together was built using Cstacks; individual genotypes were then compared to the catalog using sstacks, and rxstacks was used to exclude problematic loci. Finally, we used the POPULATIONS program[56,57] to identify the loci present in at least 85% of individuals (-r 0.85), with a minimum stack depth of 10 (-m 10), a Minor Allele Frequency (MAF) of 1% (–min_maf 0.01), and ddRAD tags requested to be present in all populations (-p 16). The final analysis included only the first SNP per locus.

Quality control of the genomic data.   For all populations, we determined the number of raw sequence reads and unlinked SNPs. We also assessed deviation from Hardy–Weinberg equilibrium (HWE) using the exact test, which is based on Monte Carlo permutations of alleles and is the most appropriate for small sample sizes[59]. HWE tests were done using the adegenet package[60,61] implemented in R[60,61]. We used the Genepop 4.0[62] program to test for linkage disequilibrium (LD) between loci in each population, calculating exact probabilities with a Markov Chain consisting of 100 batches and 5000 iterations per batch. After adjusting the *p* value, SNPs that failed the HWE test and SNP pairs in LD in at least seven locations (corrected for multiple k tests using the sequential Bonferroni procedure[63]) were excluded from further analyses. Based on the final dataset, we calculated minor allele frequencies for *B. oxycarpa* using the adegenet[60,61] package in R[64].

Detection of outlier loci.   We used the Bayescan software to identify SNP loci as having higher (divergent selection) or lower (balancing selection) levels of population divergence than strictly neutral loci[65]. This software incorporates locus and population-specific regression terms, avoiding unrealistic assumptions such as island migration models, symmetrical gene flow, or equal population sizes[65,66]. Bayescan was run with 20 pilot runs of 10,000 iterations, a burn-in of 50,000, and a final run of 100,000 iterations. We set odds of the neutral model to 10,000 (i.e., the neutral model is 10,000 times more likely than the model with selection[65]) to minimize false-positives.

Population genomic analyses.   To investigate the effects of the river on *B. oxycarpa* population structure, we assessed the genetic structure and the historical connectivity patterns between populations along and across the river using complementary genetic analyses. The methods are described in Nazareno et al.[19], and a brief overview is provided below.

First, we calculated genetic distance among populations (D_A[67]) and visualized the results by applying multidimensional scaling (MDS) in XL-STAT (Addinsoft). This was done with the Scaling by MAjorizing a COnvex Function (SMACOF) method. As an ordination technique, MDS plots populations with similar genetic structure closer together in ordination space as established by a stress factor. No assumptions relating to the cause of structure, HWE, or gametic equilibrium are required. We used the GENELAND 4.0.2[68] package in R to develop a Bayesian model to better understand the geographic distribution of genetic variability. This approach incorporates spatial data while identifying spatially explicit genetic discontinuities, minimizing the Hardy-Weinberg and linkage disequilibrium that would result if individuals from different, randomly mating populations were incorrectly grouped. The spatial model with correlated allele frequencies proposed by Guillot et al.[68,69] was used as it enables the inference of differentiation due to limited gene flow caused by physical barriers. We conducted 100 independent runs of 1,000,000 in length, discarding the burn-in of 500,000 iterations in post-processing. As the most likely number of k populations was unknown, it was treated as a simulated variable along with the MCMC simulations ($1 \leq k \leq 16$). The modal number of genetic groups of the best run (based on posterior density values) was considered as the number of genetic clusters (K).

We used ANOVA to assess pairwise genetic differentiation (F_{ST}) following Weir & Cockerham[70] and SPAGeDi[71] to calculate F_{ST}. The significance of deviation of F_{ST} values were estimated using a jackknife procedure across loci. To assess visual similarity between genetic and geographic distances based on both the MDS and GENELAND methods, we used a Mantel test[72] for isolation by distance (IBD) to verify if the overall pattern met the expectation of decreasing genetic similarity with increasing geographic distance. To test the riverine barrier hypothesis for the Rio Negro, we deconstructed the genetic structure using a multiple matrix regression, allowing us to assess the relative contribution of long-term historical divergence and the effects of IBD. The model proposed by Legendre & Legendre[73] was used as it evaluates the relationship across three matrices: (1) pairwise genetic distances [$F_{ST}/(1 - F_{ST})$[74]] between *B. oxycarpa* populations; (2) Euclidian distances representing the geographic distance between pairwise *B. oxycarpa* populations; and (3) a pairwise binary matrix of isolation by the river as an expression of long-term historical divergence. This binary matrix was constructed by coding each *B. oxycarpa* population pair in relation to the river, with populations on the same river bank as '0' and those on different river banks as '1'[27]. Multiple matrix regression and a single Mantel test were performed in R[64] using 10,000 permutation tests of significance for the correlation coefficient.

To examine the effect of the river on genetic variation between populations, we used a nested hierarchical analysis of molecular variance (AMOVA[75]), defining two hierarchical levels of population differentiation: between populations from opposite river banks; and between populations along each bank. Arlequin 3.5.2[76] was used to calculate population differentiation estimates and their statistical significance based on 20,000 random permutations.

To quantify historical connectivity based on neutral loci, we used TreeMix 1.12[77] to construct historical relationships between populations based on a population graph analysis that permits population divergence and migration. Because the model used in Treemix allows for population differentiation in the presence of post-divergence admixture/migration (m), it improves the likelihood fit of a bifurcating phylogeny. The branch lengths of the resulting phylogeny are proportional to the amount of genetic drift per branch, based on a composite maximum likelihood of the local optimum tree[78]. Thus, inference is based on "shared genetic drift" between sets of populations, assuming that shared drift implies a shared evolutionary history[79]. We added stepwise migration edges, inspecting the results for consistency between runs, and we used R[64] to visualize the population graph and residuals.

Data availability
SNP dataset is available for download from the Dryad Digital Repository (DOI: https://doi.org/10.5061/dryad.8pk0p2nht).

References

1. Wright, S. Isolation by distance. *Genetics* **28**, 114–138 (1943).
2. Levin, D. A. & Kerster, H. W. Gene flow in seed plants. *Evolutionary Biology* **7**, 139–220 (1974).
3. Ellstrand, N. C. Is gene flow the most important evolutionary force in plants? *American Journal of Botany* **101**, 737–753 (2014).
4. Hamrick, J. L., Murawski, D. A. & Nason, J. D. The influence of seed dispersal mechanisms on the genetic structure of tropical tree populations. *Vegetatio* **108**, 281–297 (1993).
5. Morjan, C. L. & Rieseberg, L. H. How species evolve collectively: Implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**, 1341–1356 (2004).
6. Burney, C. W. & Brumfield, R. T. Ecology predicts levels of genetic differentiation in Neotropical birds. *The American Naturalist* **174**, 358–368 (2009).
7. Brumfield, R. T. & Capparella, A. P. Historical diversification of birds in northwestern South America: A molecular perspective on the role of vicariant events. *Evolution* **50**, 1607–1624 (1996).
8. Cheviron, Z. A., Hackett, S. J. & Capparella, A. P. Complex evolutionary history of a Neotropical lowland forest bird (*Lepidothrix coronata*) and its implications for historical hypotheses of the origin of Neotropical avian diversity. *Molecular Phylogenetics and Evolution* **36**, 338–357 (2005).
9. Eberhard, J. R. & Bermingham, E. Phylogeny and comparative biogeography of *Pionopsitta* parrots and *Pteroglossus* toucans. *Molecular Phylogenetics and Evolution* **36**, 288–304 (2005).
10. Moraes, L. J. C. L., Pavan, D., Barros, M. C. & Ribas, C. C. The combined influence of riverine barriers and flooding gradients on biogeographical patterns for amphibians and squamates in south-eastern Amazonia. *Journal of Biogeography* **43**, 2113–2124 (2016).
11. Storfer, A. *et al*. Putting the 'landscape' in landscape genetics. *Heredity* **98**, 128–142 (2007).
12. Holderegger, R., Buehler, D., Gugerli, F. & Manel, S. Landscape genetics of plants. *Trends in Plant Science* **15**, 675–683 (2010).
13. Abbasi, S., Afsharzadeh, S., Saeidi, H. & Triest, L. Strong genetic differentiation of submerged plant populations across mountain ranges: Evidence from *Potamogeton pectinatus* in Iran. *PLoS ONE* **11**(8), e0161889, https://doi.org/10.1371/journal.pone.0161889 (2016).
14. Rachmat, H. H., Susilowati, A., Elfiati, D., Hartini, K. S. & Faradillah, W. N. Strong genetic differentiation of the endemic rosin-producing tree *Styrax sumatrana* (Styracaceae) in North Sumatra, Indonesia. *Biodiversitas* **18**, 1331–1335 (2017).
15. Jinga, P. & Ashley, M. V. A mountain range is a strong genetic barrier between populations of *Afzelia quanzensis* (pod mahogany) with low genetic diversity. *Tree Genetics & Genomes* **14**, https://doi.org/10.1007/s11295-017-1217-x (2018).
16. Hoesen, D. S. H. Styrax L. (eds. Boer, E. & Ella, A. B.) 112–119. (Backhuys Publishers, 2000).
17. Collevatti, R. G., Leoi, L. C. T., Leite, S. A. & Gribel, R. Contrasting patterns of genetic structure in *Caryocar* (Caryocaraceae) congeners from flooded and upland Amazonian forests. *Biological Journal of the Linnean Society* **98**, 278–290 (2009).
18. Nazareno, A. G., Dick, C. W. & Lohmann, L. G. Wide but not impermeable: Testing the riverine barrier hypothesis for an Amazonian plant species. *Molecular Ecology* **26**, 3636–3648 (2017a).
19. Nazareno, A. G., Dick, C. W. & Lohmann, L. G. Tangled banks: a landscape genomic evaluation of wallace's riverine barrier hypothesis for three amazon plant species. *Molecular Ecology* **5**, 980–997 (2018).
20. Wallace, A. R. On the monkeys of the Amazon. *The Annals and Magazine of Natural History* **14**, 451–454 (1854).
21. Sosa, P. A., González-Gonzáles, E. A., González-Pérez, M. A. & De Paz, P. L. P. Contrasting patterns of genetic differentiation in Macaronesian lineages of *Ilex* (Aquifoliaceae). *Botanical Journal of the Linnean Society* **173**, 258–268 (2013).
22. Marquete, N. F. S. & Valente, M. C. *Combretaceae*. (ed. Forzza, R.C.) 864–866 (Jardim Botânico do Rio de Janeiro, 2010).
23. Neto, R. L. S. N., Cordeiro, L. S. & Loiola, I. B. Flora do Ceará, Brasil: Combretaceae. *Rodriguésia* **65**, 685–700 (2014).
24. Thien, L. B. Floral biology of *Magnolia*. *American Journal of Botany* **61**, 1037–1045 (1974).
25. Knogge, C. & Heymann, E. W. Seed dispersal by sympatric *Tamarins*, *Saguinus mystax* and *Saguinus fuscicollis*: Diversity and characteristics of plant species. *Folia Primatologica* **74**, 33–47 (2003).
26. Kubitzki, K. & Ziburski, A. Seed dispersal in flood plain forests of Amazonia. *Biotropica* **26**, 30–43 (1994).
27. Nazareno, A. G., Bemmels, J. B., Dick, C. W. & Lohmann, L. G. Minimum sample sizes for population genomics: An empirical study from an Amazonian plant species. *Molecular Ecology Resources* **17**, 1136–1147 (2017b).
28. Naka, L. N., Bechtoldt, C. L., Henriques, L. M. P. & Brumfield, R. T. The role of physical barriers in the location of avian suture zones in the Guiana Shield, northern Amazonia. *The American Naturalist* **179**, E115–E132 (2012).
29. Ribas, C. C., Aleixo, A., Nogueira, A. C. R., Miyaki, C. Y. & Cracraft, J. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proceedings of the Royal Society of London Series B-Biological Sciences* **279**, 681–689 (2012).
30. Godinho, M. B. C. & da Silva, F. R. The influence of riverine barriers, climate, and topography on the biogeographic regionalization of Amazonian anurans. *Scientific Reports* **8**, 3427 (2018).
31. Boubli, J. P. *et al*. Spatial and temporal patterns of diversification on the Amazon: A test of the riverine hypothesis for all diurnal primates of Rio Negro and Rio Branco in Brazil. *Molecular Phylogenetics and Evolution* **82**, 400–412 (2015).
32. Fiaschi, P. & Pirani, J. R. Review of plant biogeographic studies in Brazil. *Journal of Systematics and Evolution* **47**, 477–496 (2009).
33. Santorelli, S., Magnusson, W. E. & Deus, C. P. Most species are not limited by an Amazonian river postulated to be a border between endemism areas. *Scientific Reports* **8**, 2294 (2018).
34. Baker, P. A. *et al*. The emerging field of geogenomics: Constraining geological problems with genetic data. *Earth Science Reviews* **135**, 38–47 (2014).
35. Dick, C. W., Hardy, O. J., Jones, F. A. & Petit, R. J. Spatial scales of pollen and seed-mediated gene flow in tropical rain forest trees. *Tropical Plant Biology* **1**, 20–33 (2008).
36. Phillipsen, I. C. & Lytle, D. A. Aquatic insects in a sea of desert: Population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. *Ecography* **36**, 731–743 (2013).
37. Hopken, M. W., Douglas, M. R. & Douglas, M. E. Stream hierarchy defines riverscape genetics of a North American desert fish. *Molecular Ecology* **22**, 956–971 (2013).
38. Hyslop, J. & Trowsdale, S. A review of hydrochory (seed dispersal by water) with implications for riparian rehabilitation. *Journal of Hydrology* **51**, 137–152 (2012).
39. Boyle, S. A. *et al*. Geographic comparison of plant genera used in frugivory among the Pitheciids *Cacajao*, *Callicebus*, *Chiropotes*, and *Pithecia*. *American Journal of Primatology* **78**, 493–506 (2016).
40. Gentry, A. H. Patterns of neotropical plant diversity. *Evolutionary Biology* **15**, 1–84 (1982).
41. Peres, C. A. & van Roosmalen, M. *Primate frugivory in two species-rich neotropical forests: Implications for the demography of large-seeded plants in overhunted areas*. (eds. Levey, D. J., Silva, W. R. & Galetti, M.) 407–421 (CABI Publishing, 2002).
42. Stewart, D. *Rio Negro, rich life in poor water*. (eds. Goulding, M., Carvalho, M.L. & Ferreira, E.G.) 514–514 (SPB Academic Publishing 1989).
43. Ayres, J. M. & Clutton-Brock, T. H. River boundaries and species range size in Amazonian primates. *The American Naturalist* **140**, 531–537 (1992).
44. Gascon, C. *et al*. Riverine barriers and the geographic distribution of Amazonian species. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 13672–13677 (2000).
45. Latrubesse, E. M., Stevaux, J. C. & Sinha, R. Tropical rivers. *Geomorphology* **70**, 187–206 (2005).

46. Almeida-Filho, R. & Miranda, F. P. Mega capture of the Rio Negro and formation of the Anavilhanas Archipelago, Central Amazonia, Brazil: Evidences in an SRTM digital elevation model. *Remote Sensing of Environment* **110**, 387–392 (2007).
47. Ferreira, L. V. Effect of flooding duration on species richness, floristic composition and forest structure in river margin habitats in Amazonian blackwater floodplain forests: Implications for future design of protected areas. *Biodiversity Conservation* **9**, 1–14 (2000).
48. Keel, S. H. & Prance, G. T. Studies of the vegetation of a black water igapó (Rio Negro-Brazil). *Acta Amazônica* **9**, 645–655 (1979).
49. Macedo, M. & Prance, G. T. Notes on the vegetation of Amazonia II. The dispersal of plants in Amazonian white sand campinas: The campinas as functional islands. *Brittonia* **30**, 203–215 (1978).
50. Brown, J. E. *et al*. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. *Evolution* **68**, 514–525 (2014).
51. Puckett, E. E. *et al*. Global population divergence and admixture of the brown rat (*Rattus norvegicus*). *Proceedings of the Royal Society B: Biological Sciences* **283**, 20161762 (2016).
52. Trucchi, E. *et al*. Long live the alien: Is high genetic diversity a pivotal aspect of crested porcupine (*Hystrix cristala*) long-lasting and successful invasion? *Molecular Ecology* **25**, 3527–3539 (2016).
53. Kotsakiozi, P. *et al*. Population genomics of the Asian tiger mosquito, *Aedes albopictus*: insights into the recent worldwide invasion. *Ecology and Evolution* **7**(23), 10143–10157 (2017).
54. Willing, E.-M., Dreyer, C. & van Oosterhout, C. Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS One* **7**(**8**), https://doi.org/10.1371/journal.pone.0042649 (2012).
55. Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S. & Hoekstra, H. E. Double digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135 (2012).
56. Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. Stacks: Building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* **1**, 171–182 (2011).
57. Catchen, J. M., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. Stacks: An analysis tool set for population genomics. *Molecular Ecology* **22**, 3124–3140 (2013).
58. Hohenlohe, P. A. *et al*. Population genomics of parallel adaptation in three-spine stickleback using sequenced RAD tags. *PLoS Genetics* **6**, e1000862 (2010).
59. Barnholtz-Sloan, J. S. *Population Genetics*. (eds Krawetz, S. A. & Womble, D. D.) (Humana Press, 2003).
60. Jombart, T. Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).
61. Jombart, T. & Ahmed, I. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**, 3070–3071 (2011).
62. Rousset, F. Genepop'007: A complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology Resources* **8**, 103–106 (2008).
63. Rice, W. R. Analyzing tables of statistical tests. *Evolution* **43**, 223–225 (1989).
64. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (2015).
65. Foll, M. & Gaggiotti, O. E. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993 (2008).
66. Nielsen, E. E. *et al*. Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**, 276 (2009).
67. Nei, M., Tajima, F. & Tateno, Y. Accuracy of estimated phylo-genetic trees from molecular data. *Journal of Molecular Evolution* **19**, 153–170 (1983).
68. Guillot, G. Population genetic and morphometric data analysis using R and the Geneland program, https://www2.imm.dtu.dk/~gigu/Geneland/Geneland-Doc.pdf (2012).
69. Guillot, G., Santos, F. & Estoup, A. Analyzing georeferenced population genetics data with Geneland: A new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics* **24**, 1406–1407 (2008).
70. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
71. Hardy, O. J. & Vekemans, X. SPAGeDi: A versatile computer program to analyze spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* **2**, 618–620 (2002).
72. Mantel, N. The detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209–220 (1967).
73. Legendre, P. & Legendre, L. *Numerical Ecology*. (Elsevier Science, 1998).
74. Rousset, F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* **145**, 1219–1228 (1997).
75. Excoffier, L., Smouse, P. & Quattro, J. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* **131**, 479–491 (1992).
76. Excoffier, L. & Lischer, H. E. L. 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 (2010).
77. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics* **8**(11), e1002967 (2012).
78. Felsenstein, J. Evolutionary trees from gene frequencies and quantitative characters: Finding maximum likelihood estimates. *Evolution* **35**, 1229–1242 (1981).
79. Peter, B. M. Admixture, population structure, and F-statistics. *Genetics* **202**, 1485–1501 (2016).

Acknowledgements

Author contributions

A.G.N. and L.G.L. designed the study and coordinated sample collection. A.G.N. conducted molecular work, performed analyses, and led the writing of the manuscript with input from all co-authors. C.W.D. provided laboratory assistance, analytical input and troubleshooting.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-55147-1.

Correspondence and requests for materials should be addressed to A.G.N. or L.G.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.