










# Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity

Rowan J. Schley<sup>1,2</sup>  | R. Toby Pennington<sup>3,4</sup>  | Oscar Alejandro Pérez-Escobar<sup>1</sup>  | Andrew J. Helmstetter<sup>5</sup>  | Manuel de la Estrella<sup>6</sup>  | Isabel Larridon<sup>1,7</sup>  | Izai Alberto Bruno Sabino Kikuchi<sup>1,8</sup> | Timothy G. Barraclough<sup>2,9</sup>  | Félix Forest<sup>1</sup>  | Bente Klitgård<sup>1</sup> 

<sup>1</sup>Royal Botanic Gardens, Kew, Richmond, UK

<sup>2</sup>Department of Life Sciences, Imperial College London, Ascot, Berkshire, London, UK

<sup>3</sup>Geography, University of Exeter, Exeter, UK

<sup>4</sup>Royal Botanic Garden Edinburgh, Edinburgh, UK

<sup>5</sup>Institut de Recherche pour le Développement (IRD), UMR-DIADE, Montpellier, France

<sup>6</sup>Departamento de Botánica, Ecología y Fisiología Vegetal, Facultad de Ciencias, Universidad de Córdoba, Córdoba, Spain

<sup>7</sup>Systematic and Evolutionary Botany Lab, Department of Biology, Ghent University, K.L., Gent, Belgium

<sup>8</sup>Universiteit Leiden, Hortus botanicus Leiden, Leiden, The Netherlands

<sup>9</sup>Department of Zoology, University of Oxford, Oxford, UK

## Correspondence

Rowan J. Schley, Room E.2.6, Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK.

Email: [r.schley@kew.org](mailto:r.schley@kew.org)

## Funding information

Genetics Society; Natural Environment Research Council

## Abstract

Hybridization has the potential to generate or homogenize biodiversity and is a particularly common phenomenon in plants, with an estimated 25% of plant species undergoing interspecific gene flow. However, hybridization in Amazonia's megadiverse tree flora was assumed to be extremely rare despite extensive sympatry between closely related species, and its role in diversification remains enigmatic because it has not yet been examined empirically. Using members of a dominant Amazonian tree family (*Brownea*, Fabaceae) as a model to address this knowledge gap, our study recovered extensive evidence of hybridization among multiple lineages across phylogenetic scales. More specifically, using targeted sequence capture our results uncovered several historical introgression events between *Brownea* lineages and indicated that gene tree incongruence in *Brownea* is best explained by reticulation, rather than solely by incomplete lineage sorting. Furthermore, investigation of recent hybridization using ~19,000 ddRAD loci recovered a high degree of shared variation between two *Brownea* species that co-occur in the Ecuadorian Amazon. Our analyses also showed that these sympatric lineages exhibit homogeneous rates of introgression among loci relative to the genome-wide average, implying a lack of selection against hybrid genotypes and persistent hybridization. Our results demonstrate that gene flow between multiple Amazonian tree species has occurred across temporal scales, and contrasts with the prevailing view of hybridization's rarity in Amazonia. Overall, our results provide novel evidence that reticulate evolution influenced diversification in part of the Amazonian tree flora, which is the most diverse on Earth.

## KEYWORDS

Angiosperms, hybridization, phylogenomics, population genomics, rainforest, speciation

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. Molecular Ecology published by John Wiley & Sons Ltd

## 1 | INTRODUCTION

Reproductive isolation is often seen as a prerequisite for speciation and as a defining feature of species (Barraclough, 2019; Mayr, 1942). Despite this, hybridization between species is known to occur and has several different outcomes, from the erosion of evolutionary divergence (Kearns et al., 2018; Vonlanthen et al., 2012) to the formation of entirely new “hybrid species” (Mallet, 2007). Neotropical rainforests, and in particular Amazonia, harbour the highest levels of plant diversity on Earth (Cardoso et al., 2017; Ulloa Ulloa et al., 2017), the evolution of which was influenced by many drivers (Antonelli & Sanmartín, 2011). However, to date there has been little convincing evidence that hybridization consistently occurs between tree species therein. Indeed, the prevailing view has been that hybridization between tropical tree species is an exceptionally rare event (Ashton, 1969; Ehrendorfer, 1970; Gentry, 1982). Although a few observations of reproductive biology support this, such as the intersterility between Costa Rican lineages of *Inga*, a species-rich legume genus (Koptur, 1984), evidence for hybridization has been poorly tested empirically in Amazonian trees (in contrast to other Neotropical regions and taxa (e.g. Morales-Briones, Liston, & Tank, 2018)). Indeed, within Amazonia the only potential molecular evidence of hybridization among tree species involved two species of *Carapa* (Scotti-Saintagne et al., 2013), and a recent review of hybrid zones (which are one potential outcome of hybridization) identified no studies focussed on Amazonian taxa (Abbott, 2017). Since Amazonia is the largest expanse of rainforest in the world, containing at least 6,800 tree species (Cardoso et al., 2017) and given that hybridization is widely acknowledged as a powerful creative force in plant evolution, being known to generate morphological and genetic novelty (Rieseberg et al., 2007), the paucity of studies examining it as an aspect of evolution in the hyperdiverse rainforests of Amazonia is surprising.

One factor which may allow more widespread hybridization in Amazonian tree species than was previously assumed, if their reproductive isolation is not absolute, is the remarkable level of sympatry found for closely related species. In several Amazonian lineages, including *Inga*, *Gutteria* and *Swartzia*, many recently diverged species co-occur (Dexter et al., 2017), often to a remarkable degree. One such example of this is the co-occurrence of 19 *Inga* species in a single hectare of the Ecuadorian Amazon (Valencia, Balslev, & Miño, 1994). As such, for many rainforest taxa the opportunity for hybridization is constantly present.

Hybridization can have a range of evolutionary consequences. In many cases, hybridization simply results in the formation of sterile, maladapted offspring with poor reproductive fitness, with subsequent selection against them maintaining isolation between species (such as in reinforcement, where selection against hybrids drives the evolution of prezygotic isolation (Hopkins, 2013)). In other cases, there may be a significant movement of genetic material from one lineage to another facilitated by backcrossing (Rieseberg & Wendel, 1993), which is known as “introgression.” This transfer of genetic material through hybridization may confer a selective

advantage to resultant offspring (Taylor & Larson, 2019), in which case it is referred to as “adaptive introgression.” Adaptive introgression is often observed between closely related taxa during the invasion of new habitats (e.g. Suarez-Gonzalez, Lexer, & Cronk, 2018; e.g. Whitney, Randell, & Rieseberg, 2010). Furthermore, hybridization can lead to rapid evolutionary radiations. This occurs through the re-assembly of standing genetic variation which has accumulated between diverging lineages, and which has already been subject to selection. This “combinatorial” process is much more rapid than the gradual accumulation of variation through mutation, and the passage of these variants through hybridization often fuels rapid diversification events (Marques, Meier, & Seehausen, 2019). This has so far been demonstrated largely in animals, most notably in African cichlid fish (Meier et al., 2017).

Introgression can occur at different rates in different regions of the genome (Payseur, 2010). Regions under divergent selection may remain distinct due to reduced fitness of hybrid genotypes at such loci, resulting in a low rate of introgression, as demonstrated in temperate-zone tree species, where divergence between hybridizing lineages is maintained through selection (Hamilton & Aitken, 2013; Sullivan, Owusu, Weber, Hipp, & Gailing, 2016). This explains why species that hybridize can remain as biologically distinct (and taxonomically identifiable) groups despite undergoing genetic exchange with other species (Abbott et al., 2013; Seehausen et al., 2014). Conversely, regions under little or no selection tend to introgress more freely, becoming homogenized between species. Moreover, if there is selection for hybrid genotypes (as in adaptive introgression), the rate of introgression for a region may be further increased relative to the rest of the genome (Gompert, Parchman, & Buerkle, 2012).

*Brownea*, a member of the legume family (Fabaceae), is a characteristic tree genus of lowland Neotropical rainforests and contains around 27 species distributed across northern South America, particularly within Amazonia. Previous work indicates that there is a broad degree of phylogenetic incongruence evident in this genus (Schley et al., 2018) which might indicate hybridization, although this could also result from incomplete lineage sorting. However, there are numerous *Brownea* hybrids in cultivation (e.g. *B. x crawfordii* (Crawford & Nelson, 1979) and *B. hybrida* (Backer, 1911)), as well as several instances of putative hybridization among *Brownea* lineages in the wild. The most notable of these instances is that proposed between the range-restricted *Brownea jaramilloi* (Pérez, Klitgård, Saslis-Lagoudakis, & Valencia, 2013) and the wide-ranging *B. grandiceps*, which co-occur in the Ecuadorian Amazon (Figure 1). There are multiple morphological distinctions between these two species, including differences in inflorescence colour and structure, growth habit, tree height and leaf morphology (Pérez et al., 2013). Although they co-occur, these species favour different habitats: *B. jaramilloi* grows on ridge tops and hillsides, whereas *B. grandiceps* shows a slight preference for swamps and valleys but is more evenly distributed (pers. obs. & Klitgaard, 1991; Pérez et al., 2013). Despite this, hybridization appears to occur, as evidenced by the existence of a putative hybrid between these



**FIGURE 1** Two co-occurring *Brownea* lineages (*Brownea grandiceps* (photograph © Rowan Schley) and *Brownea jaramilloi* (photograph © Xavier Cornejo)) as well as their putative hybrid *Brownea* “rosada” (photograph © J. L. Clark)

two species known as *B. “rosada”* (Figure 1). *Brownea “rosada”* displays an intermediate morphology between its two parental species, producing pink flowers. The hypothesis of a *B. jaramilloi* × *B. grandiceps* hybrid has not yet been tested, however, using molecular data.

As a member of the legume family, which dominates Amazonian forests (ter Steege et al., 2013), and with its apparent propensity for hybridization, *Brownea* is an excellent system with which to study the phylogenetic patterns and genomic architecture of introgression in Amazonian trees. Systematically documenting hybridization at a range of timescales and taxonomic levels within this group could help understand how admixture has contributed to the assembly of one of the world's richest floras. Accordingly, this study investigates the role of hybridization in the diversification of this rainforest tree genus at two levels. First, we test whether introgression played a role during the long-term diversification of the genus, by investigating reticulation at deep phylogenetic levels. Second, we test whether recent gene flow between *Brownea* species occurs at the landscape scale and if so, whether it occurs evenly across the genome or differentially for subsets of loci.

## 2 | MATERIALS AND METHODS

### 2.1 | Phylogenetic analyses

In order to test for ancient reticulation between *Brownea* species, sequences from 225 nuclear genes were used to elucidate evolutionary relationships. This was done using a targeted bait capture sequencing approach along with phylogenomic methods, for which 23 of 27 lineages were sampled within *Brownea*, including the three subspecies of *B. coccinea* and *B. “rosada,”* a putative hybrid lineage of *B. grandiceps* and *B. jaramilloi*. In total, 59 accessions were sampled within *Brownea*, as well as an additional 13 outgroup taxa from the genera *Macrobium*, *Heterostemon*, *Paloue* and *Browneopsis*, which form part of the “*Brownea* clade” (Fabaceae, subfamily Detarioideae

(LPWG, 2017)). The list of accessions and their associated information can be found in Table S1.

DNA sequence data were generated using leaf material collected from herbarium specimens and silica-gel-dried accessions. Genomic DNA was extracted using the CTAB method (Doyle & Doyle, 1987), and sequencing libraries were prepared using the NEBNext® Ultra™ II DNA Kit (New England Biolabs, Massachusetts, USA) with a modified protocol to account for fragmented DNA, and including a ~600 bp size-selection step. Targeted bait capture was performed using the MyBaits protocol (Arbor Biosciences, Michigan, USA) to target 289 phylogenetically informative nuclear genes, using a bait kit designed for the legume subfamily Detarioideae within which *Brownea* is nested (Ojeda et al., 2019). The final library pools were sequenced either on the Illumina MiSeq platform (Illumina, San Diego, USA) at RBG Kew, or on the Illumina HiSeq platform by Macrogen Inc. (Seoul, South Korea).

DNA sequencing reads were quality-checked with the program FASTQC v0.11.3 (Andrews, 2010) and were subsequently trimmed using Trimmomatic v.0.3.6 (Bolger, Lohse, & Usadel, 2014) in order to remove adapter sequences and to quality filter reads. Trimmomatic settings permitted <4 mismatches, a palindrome clip threshold of 30 and a simple clip threshold of 6. Bases with a quality score <28 and reads shorter than 36 bases long were removed from the data set. Following quality filtering, loci were assembled using SPAdes v3.11.1 (Bankevich et al., 2012) by the HybPiper pipeline v1.2 (Johnson et al., 2016) and potentially paralogous loci were removed using the Python (Python Software Foundation, 2010) script “*paralog\_investigator.py*,” distributed with the HybPiper pipeline. All sequences were aligned by gene region using MAFFT v7.215 (Katoh & Standley, 2013) and were cleaned using the *-automated1* flag in *trimal* (Capella-Gutiérrez, Silla-Martínez, & Gabaldón, 2009), which calculates the optimal method to remove sequence gaps from alignments and removes them. In order to infer gene trees for phylogenetic network analysis, the 225 recovered gene regions were further refined to include only 20 taxa, representing a single accession per lineage in *Brownea*,

using *Macrobium colombianum* as the outgroup taxon. Where applicable, accessions were chosen by comparing the sequence recovery of conspecific samples and choosing the individual with the best gene recovery. This resulted in 220 single-accession-per-lineage gene alignments. Further details of phylogenomic sampling, labwork, sequencing and data processing are detailed in Methods S1.

The 225 cleaned, full-accession gene alignments were concatenated using *amas* (Borowiec, 2016) and were subsequently used for phylogenetic inference in *RAXML HPC2* (Stamatakis, 2014). Inference was performed using 1,000 rapid bootstrap replicates and the GTRCAT model of nucleotide substitution (which is a parameter-rich model and hence is most suitable for large data sets) on the CIPRES web portal (Miller, Pfeiffer, & Schwartz, 2010), <https://www.phylo.org/>). In addition, gene trees were generated for each of the full and single-accession gene alignments using *RAXML v.8.0.26* (Stamatakis, 2014) with the same settings as above. *Macrobium colombianum* was used to root both the full and single-accession data set analyses since it was the *Macrobium* accession that was present in the most gene alignments.

Gene trees from both data sets were used to generate species trees under the multispecies coalescent model in the heuristic version of *ASTRAL v.5.6.1* (Zhang, Rabiee, Sayyari, & Mirarab, 2018) using the default parameters. Monophyly was not enforced for individuals belonging to the same species (the “-a” option in *ASTRAL*). Finally, discordance between gene trees was calculated and visualized for the full data set using the Java program *PhyParts v.0.0.1* (Smith, Moore, Brown, & Yang, 2015) (<https://bitbucket.org/blackrim/phyParts>). The pattern of incongruence between gene trees for each node was then mapped onto the *ASTRAL* species tree using the Python script *PhyPartsPieCharts v1.0* (<https://github.com/mossmatters/MJPythonNotebooks>).

## 2.2 | Inferring ancient introgression

Phylogenetic networks were inferred for 220 gene trees from the single-accession-per-lineage data set to understand whether introgression occurred over *Brownea*'s evolutionary history. Networks were inferred with the program *SNaQ!*, implemented in the *Julia v0.6.4* (Bezanson, Edelman, Karpinski, & Shah, 2017) package *PhyloNetworks v0.11.0* (Solís-Lemus, Bastide, & Ané, 2017). This program facilitates testing of whether the observed incongruence between gene trees is better explained by a model describing only incomplete lineage sorting or by a model describing reticulation. Networks were estimated by calculating quartet concordance factors (CF) for each node from the single-accession-per-lineage gene trees, since *SNaQ!* requires that each tip of the phylogenetic trees represents a single lineage. The network with the number of hybridization events ( $h$ ) best describing the data was chosen using negative log-pseudolikelihood comparison. Finally, the observed gene tree CFs were compared with those of the best-fit phylogenetic network (i.e. the model including hybridization) and those expected under a

coalescent model (i.e. a “tree-like” model which accounts for incomplete lineage sorting but not hybridization). The fit of the observed gene tree topologies to either the “network-like” or the “tree-like” model was assessed using the Tree Incongruence Checking in R (TICR) test (Stenz, Larget, Baum, & Ané, 2015). This was done with the function “*test.one.species.tree*” in the R v3.4.4 (R Development Core Team, 2013) package *PHYLOLM* (Ho & Ané, 2014). Proportions of genes contributed between lineages by hybridization events were taken from the “Gamma value” output of *PhyloNetworks*. Further details of phylogenetic network analysis are contained in Methods S1.

In order to further investigate patterns of ancient reticulation, we used the *Dtrios* function in *Dsuite* (Malinsky, 2019) to perform Patterson's *D*-statistic tests (Durand, Patterson, Reich, & Slatkin, 2011; Green et al., 2010). Patterson's *D*-statistic uses asymmetry in gene tree topologies to quantify introgression between either of two lineages which share a common ancestor (P1 and P2) and one other lineage (P3) that diverged from the common ancestor of P1 and P2 or earlier. We generated an input VCF file for *Dtrios* based on the single-accession-per-lineage data set using *SNPsites* (Page et al., 2016) and ran *Dtrios* with *Macrobium colombianum* as an outgroup for all 969 tests. In addition, we used the single-accession-per-lineage *ASTRAL* tree to inform taxon relationships for each test. We assessed significance of each test using 100 jackknife resampling runs and visualized the most conservative *D*-statistic estimates (the “D\_min” output of *Dtrios*) with the Ruby script “*plot\_d.rb*,” available from <https://github.com/mmatschiner>.

## 2.3 | Population-level introgression between *B. grandiceps* and *B. jaramilloi*

Having examined the degree of historical reticulation among *Brownea* lineages using targeted bait capture, population genomic data were generated with ddRAD sequencing and used to investigate recent introgression at a finer taxonomic scale. More specifically, the degree of shared genetic variation was visualized for individuals of the ecologically divergent species *B. grandiceps* and *B. jaramilloi* focussing on individuals which co-occur in Yasuní National Park in the Ecuadorian Amazon, as this area appears to be a hybrid zone for these two taxa. Following this, to make inferences about the potential evolutionary significance of recent introgression the rates at which different loci introgress relative to the rest of the genome were estimated for these taxa using genomic clines.

One hundred and seventy-one specimens in total were genotyped using ddRADseq (Peterson, Weber, Kay, Fisher, & Hoekstra, 2012). Sampling consisted of 128 individuals of *Brownea grandiceps*, 40 individuals of *B. jaramilloi* and three individuals of *B. “rosada”* (the putative hybrid of *B. grandiceps* and *B. jaramilloi*), representing the relative abundance of each species in the forest plot from which they were sampled. Leaf material of sympatric *B. grandiceps* and *B. jaramilloi* trees was collected from the Yasuní National Park 50-ha forest plot in Napo, Ecuador, and dried in a herbarium press. Samples were identified based on the characters listed in Table S2, with these

morphological differences between *B. grandiceps* and *B. jaramilloi* displayed in Figure S1. The sample list is shown in Table S3.

Genomic DNA was extracted from dried leaf material with the CTAB method and purified using a Qiagen Plant Minikit (Qiagen, Hilden, Germany) column cleaning stage. Sequencing libraries were prepared by digesting the DNA template using the restriction enzymes *EcoRI* and *mspI* (New England BioLabs, Massachusetts, USA), in accordance with the ddRADseq protocol. Samples were then ligated to universal Illumina P2 adapters and barcoded using 48 unique Illumina P1 adapters. Samples were pooled and size-selected to between 375 and 550 bp using a Pippin Prep Electrophoresis machine (Sage Science, Massachusetts, USA). Following amplification and multiplexing, sequencing was performed using a single-lane, paired-end 150bp run on the HiSeq 3/4000 platform.

DNA sequencing reads from the ddRADseq genotyping were processed de novo (i.e. without the use of a reference genome) using the Stacks pipeline v2.1 (Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011). Reads were quality-filtered using Stacks by removing reads which were of poor quality (i.e. had a Phred score <10), following which “stacks” of reads were created, and SNPs were identified among all de novo “loci” and across individuals. This was done using a minimum coverage depth (the “-m” flag in Stacks) of three and a within-individual mismatch distance (-M) of seven nucleotides. Individuals with sequencing coverage under 7.5x were removed. Loci found in fewer than 40% of individuals and sites with a minor allele frequency threshold of 5% were filtered out using the “populations” module of Stacks to account for genotyping error. Model parameters for the various programs within the Stacks pipeline were chosen using the recommendations in Paris, Stevens, and Catchen (2017), as well as through running the pipeline with multiple different parameter settings. This resulted in a data set containing 22,046 loci with 120,085 SNPs for 171 individuals. A data set containing only one SNP per locus was also extracted using the Stacks *populations* module for use in analyses which assumed no linkage disequilibrium. This subsetting resulted in a data set containing 19,130 loci with 19,130 SNPs for 171 individuals. Details of population genomic sampling, library preparation, sequencing and data filtering are shown in Methods S2.

In order to understand the patterns of introgression at the population level, the single-SNP ddRAD data set containing 19,130 RAD loci was used to visualize the degree of shared variation between *B. grandiceps* and *B. jaramilloi*. This was performed using principal component analysis implemented in R v3.4.4 followed by plotting with *ggplot2* (Wickham, 2016). To further understand the patterns of shared variation, we visualized relationships between individuals by inferring a neighbour net plot in *SPLITSTREE* v4.14.6 (Huson & Bryant, 2005) and inferred population structure with the program *fastSTRUCTURE* v1.0 (Raj, Stephens, & Pritchard, 2014). The number of lineages “K” (hereafter called “groups,” as is standard in *FastSTRUCTURE*) was chosen using the value which provided the largest improvement in marginal likelihood. In addition, a *fastSTRUCTURE* run incorporating 40 random individuals from each species was performed to account for any bias which may have been incurred by differences in sample size. Finally, *NEWHYBRIDS* v1.1 (Anderson & Thompson, 2002)(<https://github.com/eriqande/newhybrids>) was used to categorize individuals into different hybrid classes, using three runs on a subset of 500 randomly selected ddRAD loci due to the computational limits of the program. Runs were performed with 50,000 MCMC sweeps following 50,000 burn-in sweeps under the default parameters of the program.

The relative “rate” and “direction” of introgression for each locus between the two *Brownea* species was estimated using Bayesian estimation of genomic clines (*bgc*) v1.03 (Gompert & Buerkle, 2011). For each locus, *bgc* compares the probability of ancestry at the locus relative to an individual's genome-wide ancestry, thereby allowing it to estimate two parameters for each locus. These parameters are  $\alpha$ , which approximately equates to the “direction” of introgression, and  $\beta$ , which may be summarized as the “rate” of introgression for a locus (Gompert & Buerkle, 2011; Gompert, Parchman, et al., 2012). In order to estimate these parameters from the single-SNP data set consisting of 19,130 RAD loci, 50,000 MCMC generations with a 50% burn-in were performed in *bgc*. Admixture proportions (i.e. mean Q values) generated by *fastSTRUCTURE* were used to assign each individual to one of three groups (*Brownea grandiceps*, *B. jaramilloi* and admixed), resulting in an admixed population containing 27 individuals. In addition, hybrid index and interspecific heterozygosity were estimated with *bgc* using the -p and -i flags, respectively, and were plotted using the *triangle.plot* function in the R package *Introgress* (Gompert & Buerkle, 2010).

Convergence was checked for the MCMC output from *bgc* in *Tracer* v1.6 (Rambaut, Suchard, Xie, & Drummond, 2015) and with the R package *coda* (Plummer, Best, Cowles, & Vines, 2006) using Geweke's diagnostic (Geweke, 1991). Loci with significant “excess ancestry” were identified by ascertaining whether the 99% posterior probability estimates of the  $\alpha$  and  $\beta$  parameters included zero (i.e. by identifying positive or negative nonzero estimates of the parameters). In addition, loci which were extreme “introgression outliers” were identified for both parameters by extracting loci whose median estimates were not included in the 99% posterior probability credible intervals (Gompert & Buerkle, 2011).

Finally, to assess whether our *bgc* analysis was limited in its ability to detect introgression outliers due to sample size, we ran an extra set of analyses based on a randomly subsampled data set consisting of 9,430 RADseq loci. These data were taken from a similar plant speciation study (Royer, Streisfeld, & Smith, 2017) and were refined to contain the same number of admixed individuals as our study ( $n = 27$ ). We then ran the *bgc* analyses with the parameters and methods described above. Further details of visualizing shared variation, hybrid category assignment and *bgc* analysis are shown in Methods S2.

Finally, to assess whether our *bgc* analysis was limited in its ability to detect introgression outliers due to sample size, we ran an extra set of analyses based on a randomly subsampled data set consisting of 9,430 RADseq loci. These data were taken from a similar plant speciation study (Royer, Streisfeld, & Smith, 2017) and were refined to contain the same number of admixed individuals as our study ( $n = 27$ ). We then ran the *bgc* analyses with the parameters and methods described above. Further details of visualizing shared variation, hybrid category assignment and *bgc* analysis are shown in Methods S2.

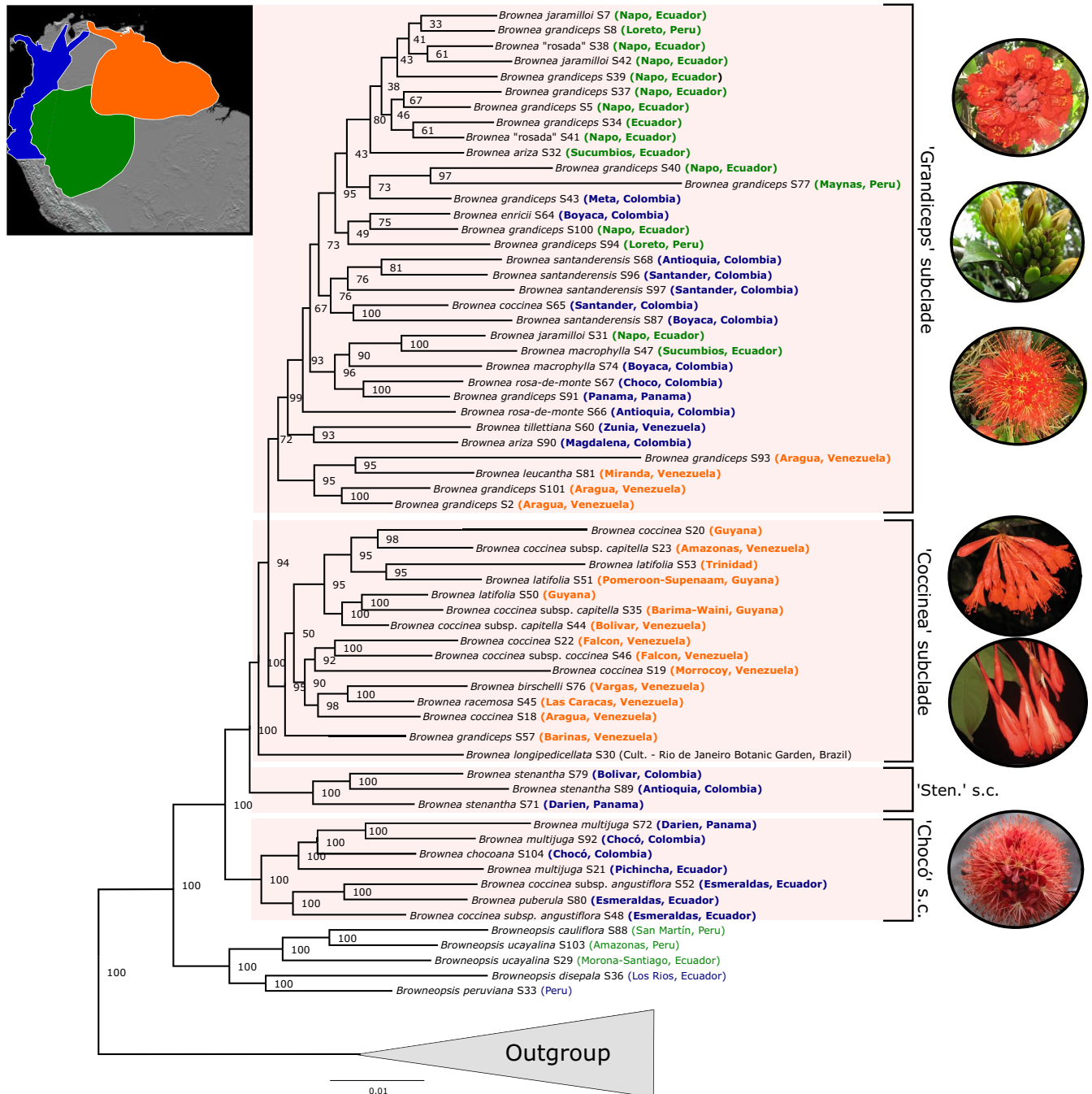
## 3 | RESULTS

### 3.1 | Phylogenetic analyses

Using DNA sequence data from 225 nuclear gene regions, we produced a concatenated phylogenetic tree using RAxML (Figure 2). This

resulted in well-supported relationships between most bipartitions (>90% bootstrap support (BS)), with lower support for some interspecific relationships, especially within the “Grandiceps” subclade. This tree also indicated a very strong signal of geographical structure

among species, with several wide-ranging species (such as *Brownea grandiceps*) being polyphyletic and being more closely related to accessions from the same region than to conspecifics. The ASTRAL tree run using the same data set resulted in a nearly identical topology



**FIGURE 2** RAXML phylogenetic tree inferred from 220 concatenated genes. Numbers at nodes of the tree signify bootstrap support (BS). Taxa within the top box belong to the “Grandiceps” subclade, taxa within the next box down belong to the “Coccinea” subclade, taxa within the third box down represent the “Stenantha” subclade, and those within the bottom box belong to the “Chocóan” subclade. The province and country from which each accession was collected are noted in parentheses next to each tip label, and colours correspond to the inset map (top left). Green corresponds to Western Amazonia, blue is the Chocó-Darién region and Andean cordilleras, while orange is Eastern Amazonia and the Guiana shield (modified from the phytogeographical regions identified by Gentry (1982)). Images show inflorescences of species within *Brownea*- species shown are, from top to bottom: *B. grandiceps* (photograph © Rowan Schley), *B. jaramilloi* (photograph © Xavier Cornejo), *B. macrophylla* (photograph © Bente Klitgaard), *B. coccinea* subsp. *capitella* (photograph © Xavier Cornejo), *B. longipedicellata* (photograph © Domingos Cardoso) and *B. multijuga* (photograph © Bente Klitgaard)

with similar levels of geographical structure and inferred a high degree of discordance among gene tree topologies at many nodes (quartet score = 0.52), with multiple alternative bipartitions reconstructed at most nodes. This is evident from the presence of many more conflicting gene trees (numbers below branches) than congruent gene trees (numbers above branches) in Figure S2.

### 3.2 | Ancient introgression in *Brownea* species

Phylogenetic networks estimated to ascertain whether diversification in *Brownea* was tree-like or reticulated pinpointed two hybridization events within the genus (Figure 3a). This is indicated by the  $-\log$ -pseudolikelihood values in Figure S3, since the number of hybridization events ( $h$ ) which best describe the data gives the largest improvement in  $-\log$  pseudolikelihood. In Figure S3,  $-\log$  pseudolikelihoods increased steadily between  $h = 0$  and  $h = 2$ , after which the increasing number of hybridization events only made minimal improvements to  $-\log$  pseudolikelihood.

The inferred phylogenetic network (Figure 3a) showed a broadly similar topology to the species tree in Figure 2, with the addition of two hybridization events between co-occurring lineages. The first hybridization event occurred between the lineage leading to the Venezuelan accessions of *B. grandiceps*/*B. leucantha* and *B. birschelli*, suggesting that the lineage leading to *B. birschelli* has in the past contributed around 34% of the genetic material present in the common ancestor of the Venezuelan *B. grandiceps* and *B. leucantha*. The second inferred occurrence of hybrid ancestry occurred between the ancestors of the subclade containing the Colombian accessions of *B. coccinea*/*B. santanderensis* and the lineage leading to *B. enricii*, which contributed around 37% of the genetic material belonging to the ancestor of the aforementioned subclade.

Tree Incongruence Checking in R (TICR) analysis indicated that a network-like model (i.e. one including reticulation) best described the patterns of incongruence in the single-accession-per-species gene trees inferred during this study. This method suggested an excess of outlier quartets ( $p = 1.240 \times 10^{-19}$ ,  $\chi^2 = 91.149$ ), which differed significantly from the CF values expected under a coalescent model describing only incomplete lineage sorting. As such, a tree-like model was able to be rejected as an explanation for the observed relationships between taxa in *Brownea*, suggesting that hybridization occurred over the course of its evolutionary history and was responsible for the observed incongruence.

Our  $D$ -statistic estimates (Figure 3b) recovered multiple significant reticulation events and indicated a pattern of geographically structured reticulation similar to that inferred by *PhyloNetworks*. More specifically, we found significant reticulation among a set of *Brownea* lineages from the Ecuadorian Amazon (*B. "rosada"* S41, *B. macrophylla* S47), another set of reticulation events between *Brownea* species from the Colombian cordilleras (*B. enricii* S64, *B. coccinea* S65, *B. rosa-de-monte* S66) and a final set of reticulation events among Venezuelan *Brownea* (*B. grandiceps* S101, *B. leucantha* S81, *B. grandiceps* S57, *B. latifolia* S50, *B. coccinea* S22). However, two

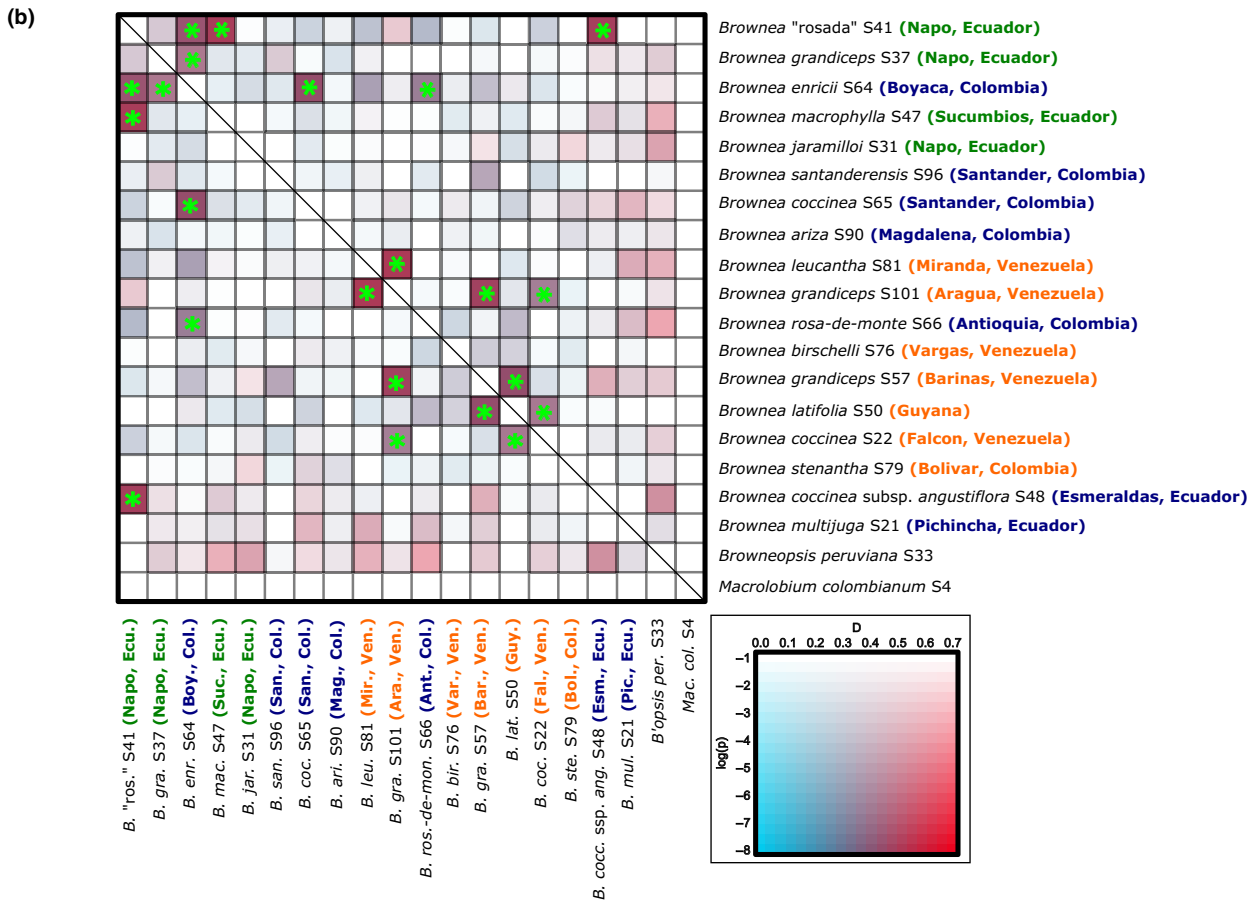
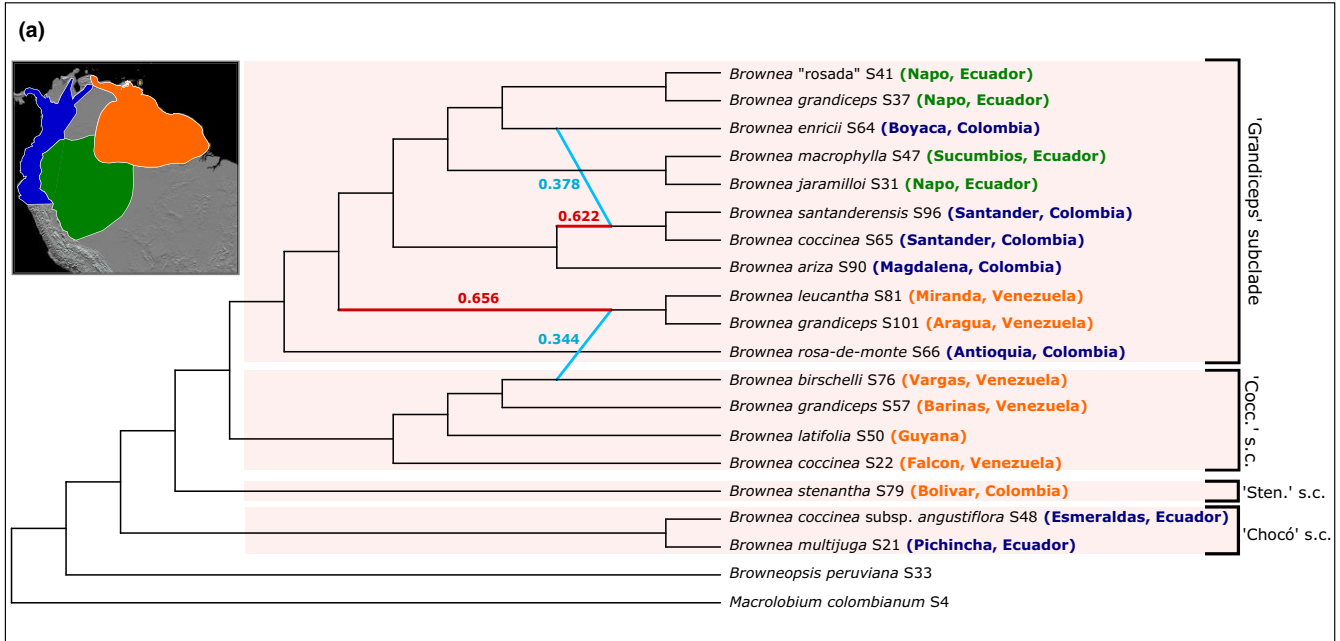
other significant reticulations were evident between *B. "rosada"* and *B. coccinea angustiflora* in Ecuador, despite being distributed either side of the Andes, as well as between Colombian *B. enricii* S64 and Ecuadorian *Brownea* (*B. "rosada"* S41 and *B. grandiceps* S37).

### 3.3 | Population-level introgression between *B. grandiceps* and *B. jaramilloi*

Our population genomic analyses revealed a broad degree of shared variation between *B. grandiceps* and *B. jaramilloi*, along with a homogeneous distribution of introgression rates across loci. In total, ddRAD sequencing resulted in 640,898,950 reads between 350 and 550bp in length for 171 individuals. Among these reads, 28,503,796 (4.45%) were discarded due to poor recovery. There was a mean of 3,747,947 reads per individual, with an average coverage depth across all samples of 27.5x. Relatively high levels of shared genetic variation were observed among taxa, along with low levels of genetic differentiation and only marginal differences in the amount of genetic variation, as shown by the population genetic statistics calculated by the Stacks pipeline for the two *Brownea* species (including *B. "rosada"*, which is grouped with *B. grandiceps*) (Table S4). The pairwise  $F_{st}$  calculated between *B. grandiceps*/*B. "rosada"* and *B. jaramilloi* was 0.111, representing a low degree of fixation, and so a high amount of shared variation.

Patterns of shared genetic variation visualized with principal component analysis (PCA), and *SPLITSTREE* showed a distinct signature of admixture between *B. grandiceps* and *B. jaramilloi*. The PCA of SNP variation inferred using the R package *ggplot2* (Figure 4a) indicated that the first principal component (PC1) explained the largest proportion of the genetic variation among all the principal components (20.4%). Individuals of *B. grandiceps* are tightly clustered along PC1, where the individuals of *B. jaramilloi* show much more variability, with many individuals forming an intergradation between the two main species clusters. Additionally, the *B. "rosada"* accessions appear to have clustered more closely to *B. grandiceps* than to *B. jaramilloi* along PC1. PC2, which explained 10.7% of the variation in the SNP data, shows a similar degree of variability in both *B. grandiceps* and *B. jaramilloi*, with two accessions of *B. "rosada"* shown to cluster in between both species. This pattern is reflected by the implicit network built using *SPLITSTREE* (Figure S4). *SPLITSTREE* recovered a clustering of individuals into two main groups, largely representing *B. grandiceps* and *B. jaramilloi*, with six putative hybrid individuals displaying an intermediate relationship between the two species clusters, and with looser clustering of morphologically identified *B. jaramilloi* individuals when compared to *B. grandiceps*.

The *fastSTRUCTURE* analysis performed to further examine the degree of shared genetic variation indicated that there was a large amount of shared ancestry between the genotyped individuals (Figure 4b), with evidence of extensive backcrossing due to the widely differing proportions of ancestry in different individuals. Marginal likelihood comparison indicated that the best value of  $K$  (i.e. the number of genetic groups) was two ( $\Delta_{\text{marginal}}$



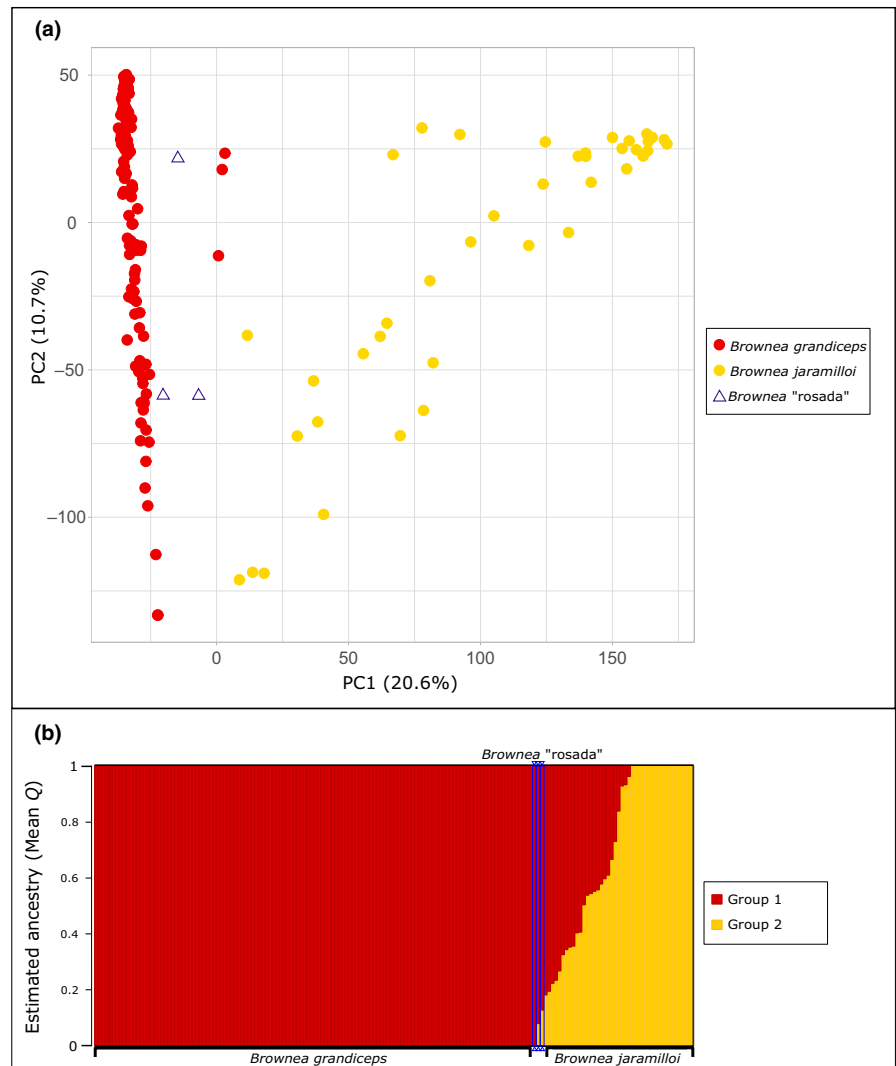
likelihood = 0.085, Figure S5a), which was the value that resulted in the largest increase in marginal likelihood. Since the "best" value of *K* is an estimate, fastSTRUCTURE plots generated with other *K* values are shown in Figure S5b). Figure 4b shows that most individuals of *B. jaramilloi* have varying fractions of *B. grandiceps* ancestry in their genome. In addition, the individuals identified as

*B. "rosada"* appear to have inherited most of their ancestry from *B. grandiceps*, with only a minimal contribution from *B. jaramilloi*. The same pattern was recovered from a fastSTRUCTURE run incorporating 40 random individuals from each species, performed to account for any bias which may have been incurred by differences in sample size (Figure S5c).



**FIGURE 3** a) Phylogenetic network with two hybridization events ( $h = 2$ ), estimated using *SNaQ!* in the Julia package *PhyloNetworks*. Light blue horizontal branches indicate inferred hybridization events, and numbers next to the branches show the estimated proportion of genes contributed by each lineage in the hybridization event. Red branches signify the ancestral lineage and what proportion of the modern lineage's genes were contributed from it. Taxa were chosen in order to represent one accession per lineage, as per the assumptions of *PhyloNetworks*, and the individual with the highest gene recovery was used for each lineage. Taxa within the top box belong to the "Grandiceps" subclade, taxa within the next box down belong to the "Coccinea" subclade, taxa within third box down represent the "Stenantha" subclade, and those within the bottom box belong to the "Chocóan" subclade. The province and country from which each accession was collected are noted in parentheses next to each tip label, and colours correspond to the inset map (top left). Green corresponds to Western Amazonia, blue is the Chocó-Darién region and Andean cordilleras, while orange is Eastern Amazonia and the Guyana shield (modified from the phytogeographical regions identified by Gentry (1982)). b) Heatmap summarizing the most conservative four-taxon D-statistic estimates and their P-values from 969 tests. Taxa P2 and P3 are displayed on the x- and y-axes in the same order as in Figure 3a, and each square represents the highest estimate of each combination of P2 and P3 taxa. The colour of each square signifies the D-statistic estimate (blue = low estimate; red = high estimate), and the colour saturation of these colours represents the P-value for that test (see inset box, bottom right). D-statistic tests for which  $p < 1 \times 10^{-6}$  (i.e. were below a Bonferroni-adjusted P-value threshold of 0.001) are marked with an asterisk (\*). As in Figure 3a, the province and country from which each accession was collected are noted in parentheses next to each tip label, and colours correspond to the inset map (top left)

**FIGURE 4** a) Principal component analysis of genotype data for all SNPs inferred at all loci. Individuals are coded by colour and point shape: red circles denote individuals identified as *B. grandiceps*, yellow circles denote individuals identified as *B. jaramilloi*, and blue triangles denote putative hybrid individuals (*B. "rosada"*). The amount of the genetic variation explained by each principal component is shown next to axes, which are labelled PC1 and PC2. b) Ancestry proportions inferred by *FastSTRUCTURE* for  $K = 2$  showing two genetic groups (group 1 in red, and group 2 in yellow). Individuals morphologically identified as the putative hybrid taxon *B. "rosada"* are labelled as such above the plot and are marked with a blue box and a blue triangle above and beneath their columns. All taxa to the left and right of these boxes were identified morphologically as *B. grandiceps* and *B. jaramilloi*, respectively, and are labelled below the x-axis



The *NEWHYBRIDS* analysis revealed that most hybrid individuals were the result of continued hybridization, with pure *B. grandiceps* making up 73.9% of the genotyped population, and pure *B. jaramilloi* making up 21.9% for the 500 subsampled loci used in this analysis. There were no F1 (first-generation) hybrids identified by the subset of loci analysed, and all hybrid individuals were either F2 hybrids (0.592%) or

had a broad distribution of probabilities across hybrid classes (3.55%), suggesting backcrossing. In the latter, the probability (Q) of samples belonging to any of the two species was below 90%; otherwise, they were identified as either *B. grandiceps* or *B. jaramilloi*.

Most patterns of shared variation inferred by *Splitstree*, *FastSTRUCTURE* and *NEWHYBRIDS* were congruent with the

morphological identification (Table S5). *Splitstree* was the most congruent (92.98% of accessions with the same classification as identified morphologically), followed by *NEWHYBRIDS* (92.31%) and *FastSTRUCTURE* (85.96%). The same hybrid class assignment was recovered for 84.62% of accessions across methods.

The Bayesian estimation of genomic clines (*bgc*) analysis, which was used to determine whether loci showed outlying rates of admixture relative to the genome-wide average, recovered a signal of asymmetric introgression, but did not detect any differential rates of introgression between loci. Of the 19,130 loci under study, *bgc* recovered 251 loci (1.3%) with positive excess ancestry for the  $\alpha$  parameter and 1,089 loci (5.69%) with negative excess ancestry for  $\alpha$ . However, no loci displayed extreme rates of introgression relative to the average rate across the genome (i.e. there were no statistically outlying  $\beta$  parameter estimates). All 27 admixed individuals had hybrid indices of between 0.4 and 0.6 and displayed relatively low levels of interspecific heterozygosity (~0.3 for all individuals) (Figure S6). The MCMC runs from which all of these results were drawn showed adequate mixing (Figure S7a-c).

The *bgc* analysis run to examine the effect of sample size on our parameter estimates using data from Royer et al. (2017) recovered multiple loci with excess ancestry for both parameters ( $\alpha = 1,439$  loci (15.26%),  $\beta = 1638$  loci (17.37%)). In addition, this analysis recovered multiple introgression outlier loci ( $\alpha = 4$  loci (0.04%),  $\beta = 15$  loci (0.16%)). All Geweke's diagnostic values for this analysis were between -1.96 and 1.96, again indicating that there was adequate MCMC mixing.

## 4 | DISCUSSION

Our study represents the first clear case of reticulate evolution among Amazonian trees documented at multiple phylogenetic levels. This contrasts with previous work, which suggested that hybridization is an extremely rare phenomenon in Amazonian trees (Ashton, 1969; Ehrendorfer, 1970; Gentry, 1982). We demonstrate that within *Brownea*, a characteristic Amazonian tree genus, reticulate evolution has occurred over the course of its evolutionary history, with evidence of hybridization deep in the history of the genus and more recently, between the closely related *B. grandiceps* and *B. jaramilloi*.

Our study recovered a low degree of support for bifurcating interspecific relationships (Figure 2) and a high degree of incongruence among gene trees (Figure S2). These results can be largely explained by the strong signal of geographically structured reticulation that we recovered, rather than being caused purely by incomplete lineage sorting, which may be common in rainforest trees (Pennington & Lavin, 2016). Indeed, the low quartet scores and minimal gene tree concordance found at the species level in this study mirror those observed in *Lachemilla*, a montane Neotropical plant genus, in which gene tree discordance was shown to be explained by both historical and recent hybridization (Morales-Briones et al., 2018), both of which we found evidence for.

### 4.1 | Reticulation has occurred at deep phylogenetic levels in *Brownea*

Phylogenetic network analysis suggested that reticulation has taken place between *Brownea* lineages in the past, with two separate hybridization events inferred (Figure 3a). The concordance factors obtained from gene trees were not adequately explained by a tree-like model (i.e. one accounting only for incomplete lineage sorting), suggesting that a network-like model (i.e. one including reticulation) best describes the diversification patterns within *Brownea* ( $P = 1.240 \times 10^{-19}$ ,  $\chi^2 = 91.149$ ). Since the inferred events of ancient introgression both involved the common ancestors of two sets of present-day species, they are likely to have occurred several million years ago, before the divergence of the descendant species. These introgression events are likely to have occurred since the Miocene period (~23Ma), as date estimates from previous work suggest that *Brownea* originated around this time (Schley et al., 2018). Thus, the detected signatures of ancient reticulation suggest that their effects in genomes have persisted over evolutionary time.

A distinct pattern of geographical structure was evident from our phylogenetic tree and network (Figure 2; Figure 3a). Wide-ranging species such as *B. grandiceps* and *B. coccinea* were polyphyletic, and co-occurring individuals from different species were more closely related to each other than to morphologically identified conspecifics. While this polyphyly could be due to stabilizing selection on certain aspects of the genome constraining morphological divergence while the rest of the genome diverges (Struck et al., 2018), this would not explain the high degree of reticulation between co-occurring species that we inferred. We found that both historical introgression events reconstructed by *PhyloNetworks* were between taxa that co-occurred (e.g. *B. grandiceps*, *B. leucantha* and *B. birschelli* in Northern Venezuela, and *B. coccinea*, *B. santanderensis* and *B. enricii* within the Colombian cordilleras), which was in agreement with our *D*-statistic tests (Figure 3b, discussed below). This geographically structured reticulation agrees with previous work (Schley et al., 2018) which indicated that the "stem" lineages within *Brownea* had a shared distribution in northern South America, perhaps facilitating hybridization.

Geographical structure is common in species complexes of tropical trees, where geography often correlates more closely with genetic affinity between related species than does morphology (Dexter, Pennington, & Cunningham, 2010). As alluded to above, this pattern can result from gene flow between sympatric, interfertile congeners which together form a syngameon (Suarez-Gonzalez et al., 2018)). Combined with the poor dispersal likely in *Brownea* given its explosively dehiscent fruit pods and tendency to form single-species stands (Klitgaard, 1991), we believe that the syngameon model best describes the reticulate evolutionary history of this group. It is also possible that our results reflect a similar model where widespread species (such as *B. grandiceps*) are progenitor lineages from which new, localized species diverge in peripatry through ecological speciation (Misiewicz and Fine, 2014; Pennington & Lavin, 2016) and subsequently hybridize in

secondary contact. However, the latter seems unlikely given that relatively distantly related, co-occurring lineages show evidence of hybridization events.

The syngameon model seems also to describe the reticulate patterns of diversification in *Brownea* when considering the results of our *D*-statistic tests (Figure 3b). These tests also recovered a geographically structured pattern of reticulation with three main centres: Ecuador, the Colombian Cordilleras and Venezuela. Geography has been shown to be a factor structuring syngameons, based simply on the increased opportunity of interbreeding when in sympatry (Boecklen, 2017). Two other significant introgression events were also inferred: one between *B. "rosada"* and *B. coccinea angustiflora* in Ecuador, and another between Colombian *B. enricii* S64 and Ecuadorian *Brownea* (*B. "rosada"* S41 and *B. grandiceps* S37). Each of these pairs of lineages has a counterpart which occurs in the Ecuadorian Amazon and another which occurs within the foothills of the Andes. As such, the inferred reticulation may be explained by hybridization between their ancestral lineages in sympatry, given the shared Andean-Amazonian distribution evident in *Brownea* (Schley et al., 2018) before the latest events of Andean orogeny around 12Ma (Hoorn et al., 2010).

Neither *PhyloNetworks* nor *D*-statistic tests recovered evidence of reticulation between *B. grandiceps*, *B. jaramilloi* and their putative hybrid *B. "rosada"*. This is despite the extensive introgression we uncovered using population genomic methods (discussed below), as well as the clustering of *B. "rosada"* with *B. grandiceps* and *B. jaramilloi* in the phylogenomic tree (Figure 2). This may be because *PhyloNetworks* is known to underestimate the amount of reticulation present in phylogenetic trees, since only certain hybridization events are considered by the program (Solís-Lemus & Ané, 2016). *PhyloNetworks* can only detect single hybridization events at each node rather than those with multiple participant lineages. In addition, hybridization events between sister lineages can only be detected if both lineages give rise to another two descendent species (Nevado, Contreras-Ortiz, Hughes, & Filatov, 2018). Moreover, it is possible that the signal of reticulation between *B. grandiceps* and *B. jaramilloi* was too recent to be favoured by both *PhyloNetworks* and the *D*-statistic tests in the presence of the deeper reticulation events we recovered.

Given the possible flaws or lack of resolution in attempting to characterize the landscape of introgression using only macroevolutionary methods such as phylogenetic networks and *D*-statistics, we examined introgression at the landscape scale using population genomic methods, as shown in the subsequent section.

#### 4.2 | Recent hybridisation also occurs between *Brownea* species, resulting in persistent hybrids and introgression across the genome

This investigation also uncovered a substantial signal of recent introgression between *B. grandiceps* and *B. jaramilloi* in the Yasuní National Park 50-ha plot located in the Ecuadorian Amazon. The

low  $F_{st}$  estimate for the *B. grandiceps*/*B. "rosada"* and *B. jaramilloi* populations (0.11), in addition to the principal component analysis (Figure 4a), the *fastSTRUCTURE* analysis (Figure 4b) and the *SPLITSTREE* plot (Figure S4), indicated a high degree of shared variation between these lineages. All these approaches show that individuals cluster into two main groups, largely representing *B. grandiceps* and *B. jaramilloi*, as well as a set of individuals forming an intergradation between the two main clusters. The individuals which form a part of this intergradation were mostly morphologically identified as *B. jaramilloi*, although hybrid individuals are also observed in the *B. grandiceps* cluster. All individuals morphologically identified as the putative hybrid lineage *B. "rosada"* had varying degrees of admixed ancestry in at least one of the three methods used to examine patterns of shared variation. As such, this suggests that *B. "rosada"* is the result of hybridization between *B. grandiceps* and *B. jaramilloi*, although none of the individuals appeared to be recent (e.g. F1) hybrids.

The higher number of variant sites and the higher nucleotide diversity ( $\pi$ ) recovered for *B. jaramilloi* (Table S4) could also reflect the hybrid ancestry of many individuals identified as this species. A similar introgression-driven increase in nucleotide diversity has been shown in closely related species of *Mimulus* which undergo asymmetric introgression (Sweigart & Willis, 2003).

The *bgc* analysis indicated that most introgression was asymmetric and predominated by *B. grandiceps*, with more loci exhibiting negative excess ancestry for the  $\alpha$  parameter (5.69% of all loci) than positive excess ancestry (1.3% of all loci). Importantly, this *bgc* analysis also indicated that gene flow occurs largely at the same rate across loci, since there were no outlying estimates of the  $\beta$  parameter. These results were apparently robust and likely unaffected by sample size, given that the analysis performed on a reduced data set from Royer et al. (2017) recovered many loci both with excess ancestry and outlying estimates for the  $\alpha$  and  $\beta$  parameters, which is concordant with what the authors recovered for their full data set.

The excess of *B. grandiceps* ancestry mirrors the asymmetry of introgression suggested by Figure 4b and is likely driven by the uneven population sizes of the two species in the Yasuní NP hybrid zone. *Brownea jaramilloi* has only ever been observed in a small part of the western Amazon (Pérez et al., 2013), whereas *B. grandiceps* occurs across northern South America, and within the Yasuní 50-ha plot *B. grandiceps* outnumbers *B. jaramilloi* by ~ 3:1. As such, the disproportionate donation of alleles from *B. grandiceps* may be due to "pollen swamping," whereby pollen transfer from another, more populous species is more likely than pollen transfer from less numerous conspecifics (Buggs & Pannell, 2006). Pollen swamping can serve as a mechanism of invasion (e.g. in *Quercus* (Petit, Bodénès, Ducouso, Roussel, & Kremer, 2004)), and the associated asymmetric introgression may lead to the extinction of the rarer species, especially when hybrids are inviable or sterile (Balao, Casimiro-Soriguer, García-Castaño, Terrab, & Talavera, 2015). However, due to the ongoing introgression evident between *B. grandiceps* and *B. jaramilloi* it appears that hybrids are not always sterile, or at least that

“foreign” alleles are not always subject to strong purifying selection at the landscape scale.

The best evidence of this lack of selection against hybrids is the absence of loci with extreme values of the  $\beta$  parameter recovered by our *bgc* analysis. Extreme deviations in  $\beta$  are mainly expected in the presence of gene flow when selection against hybrids is strong, resulting in underdominance and a paucity of heterozygous sites (Gompert, Parchman, et al., 2012). As such, our inferred lack of outlying  $\beta$  values suggests persistence of hybrid genotypes within the Yasuní 50-ha plot. Similarly, the 27 admixed individuals in this study recovered a range of hybrid indices and low interspecific heterozygosity (Fig. S6), which suggests that these individuals are the result of many generations of interspecific gene flow, resulting in “asymmetric advanced-generation introgression,” as found in previous work (De La Torre, Ingvarsson, & Aitken, 2015), although this may also result from the low *F<sub>st</sub>* values we recovered (as in Gompert, Lucas, et al., 2012).

Further to this, the observed asymmetry in introgression could be caused by adaptive introgression favouring the preferential passage of certain loci from one parental species (e.g. Yang et al., 2018). However, the lack of outlying  $\beta$  estimates for any loci renders this very unlikely, and even so, it is difficult to ascertain whether adaptive introgression has occurred without measuring the impact of this introgression on variation in the phenotype and its fitness effects (Suarez-Gonzalez et al., 2018).

The three methods used to identify lineages and hybrid classes (*Splitstree*, *FastSTRUCTURE* and *NEWHYBRIDS*) largely recovered the same identification as was made using morphology (Table S5). Between 85.96% and 92.98% of accessions had the same identification as that determined morphologically among the three methods, with 84.62% of accessions having the same identification in all three methods. The relatively small discrepancies among methods, as well as between methods when compared to the morphological identification, could have been caused by technical factors, such as through the violation of underlying assumptions of the different programs. For example, *FastSTRUCTURE* has been shown to have a slight bias for clustering by ploidy level in populations with mixed ploidy (Stift, Kolář, & Meirns, 2019), through the violation of its assumption of diploid individuals (Raj et al., 2014). While it is possible that there are allopolyploid hybrid individuals in our data set (Hegarty & Hiscock, 2009), both *B. coccinea* and *B. grandiceps* are diploid ( $2n = 24$ ) (Atchison, 1951). Moreover, the bias for clustering by ploidy level uncovered by Stift et al. (2019) was most evident when population differentiation was weak, rather than between species as in the case of our study. More generally, estimating ancestry proportions using *STRUCTURE* analysis can be affected by demographic factors (e.g. recent bottlenecks or ancient population structure) which may be the case for our data, and so, interpretation of their results should take into account that these types of analyses try to parsimoniously explain patterns of shared variation among individuals rather than explicitly test between evolutionary models (Lawson, Van Dorp, & Falush, 2018). Similarly, the fact that a reduced data set of 500 (out of a possible 19,130) SNPs was used in the *NEWHYBRIDS* analysis

means that we may have subsampled certain loci in such a way that they were not representative of some individual's genetic ancestry, giving different conservative hybrid class assignments.

The discrepancies may also have had a biological cause. For example, differential inheritance or selection on regions of the genome responsible for morphological features, when compared to the rest of the genome, may have resulted in individuals that were identified as one species morphologically, but which in actuality had hybrid ancestry (e.g. the many hybrid individuals identified as *B. jaramilloi*) (Abbott et al., 2013; Seehausen et al., 2014). Given the highly divergent phenotypes of *B. grandiceps* and *B. jaramilloi* (Figure S1; Table S2.), it is unlikely that these discrepancies were purely due to poor identification.

### 4.3 | The contribution of hybridization to Neotropical rainforest tree diversity

Studies such as ours that document persistent hybridization across evolutionary time between tree species within Amazonia, and within tropical rainforests in general, are rare. Indeed, it was previously suggested that interspecific hybrids between rainforest tree species are poor competitors and that fertile hybrid populations are nearly non-existent (Ashton, 1969). While there is some evidence of introgression in tropical trees from within Amazonia (Scotti-Saintagne et al., 2013), most available studies substantiating it are based on trees from other tropical regions or habitats (e.g. *Shorea* in Asia (Kamiya et al., 2011; Kenzo et al., 2019), or *Rhizophora* in Indo-Pacific mangroves (Lo, 2010)). Many of these other instances appear to occur only in degraded habitats or involve infertile first-generation hybrids with minimal backcrossing, which contrasts with the findings of our study. Within *Brownea*, we uncovered introgression across taxonomic, spatial and temporal scales, with evidence of backcrossing and the persistence of hybrid genotypes at the landscape scale.

It is possible that a lack of selection against hybrids allows the passage of variation between *Brownea* species, which may persist over evolutionary time and could explain why we found evidence of introgression at both macroevolutionary and microevolutionary scales. As mentioned above, this is congruent with a “syngameon” model, where large, disparate populations of individuals belonging to multiple interfertile species occasionally hybridize and exchange genetic variants. This can aid in the maintenance of genetic diversity in species which exist in very low population densities, as is typical of tropical rainforest trees, thereby preventing Allee effects (Cannon & Lerda, 2015). This has been shown in temperate tree species (Cronk & Suarez-Gonzalez, 2018), although it typically occurs at the edge of species ranges and its prevalence in hyperdiverse systems such as tropical rainforests is a topic of recent debate (e.g. Cannon & Lerda, 2019).

While it is difficult to determine whether hybridization between Amazonian tree species occurs within many syngameons of closely related lineages, or whether it is a tendency unique to certain genera such as *Brownea*, it may be prudent to review

how relationships among tropical tree lineages are inferred. Accordingly, the use of phylogenetic networks in resolving the relationships between such groups may provide additional insight into whether reticulate evolution has contributed to diversification within Amazonian rainforest, which is among the most species-rich environments on Earth (Burbrink & Gehara 2018; Solís-Lemus et al., 2017).

## ACKNOWLEDGEMENTS

Laboratory work, fieldwork and herbarium visits to NY and US were funded by the NERC SSCP DTP and a Genetics Society Heredity Fieldwork grant funded part of the fieldwork in Ecuador. Phylogenomic sequencing was also part-funded by a grant from the Royal Botanic Gardens, Kew. The authors thank Dario Ojeda for baits and associated help; Laszlo Csiba, Penny Malakasi, Niroshini Epitawalage, Dion Devey, Robyn Cowan and Steven Dodsworth for help in the laboratory; Gregor Gilfillan at the Norwegian sequencing centre for help with ddRAD sequencing; Andres Melo Burbano for his help during fieldwork; Renato Valencia and Álvaro Pérez for helping to coordinate fieldwork in Ecuador; Colin Hughes, Jeff Doyle and Matteo Fumagalli for comments on analytical methods; and Peter Raven and Benjamin Torke for their comments on the rarity of hybridization in tropical trees. Collection, transport and extraction of genetic material from plant accessions collected in Yasuní National Park were authorized by Ecuador's Environment Ministry. Permits were issued with the following authorization numbers: **Collection permit:** 021-2016-IC-FAU-FLO-DPAO-PNY; **Mobilization permit:** 037-2016-MOV-FLO-MAE-DPAO; **Export permit:** 208-2019-EXP-CM-FAU-DNB/MA; and **Genetic research permit (Contrato Marco):** MAE-DNB-CM-2018-0082.

## AUTHOR CONTRIBUTIONS

This study was conceived and all analyses were performed by R.J.S., and the study was supervised by T.G.B., F.F. and B.K. Population genomic data were generated by R.J.S., and phylogenomic data were generated by R.J.S., I.K. and I.L. Baits for hybrid capture were provided by M.d.I.E. The manuscript was written by R.J.S. with contributions from R.T.P., O.A.P.E., A.J.H., M.d.I.E., I.L., T.G.B., F.F. and B.K.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available from online repositories. All raw reads generated with the targeted bait capture and ddRADseq methods are available on the NCBI Sequence Read Archive with the Accession nos SAMN13439069-SAMN13439140 and SAMN13441804-SAMN13441974, respectively, under the BioProject number PRJNA592723. All full phylogenomic sequence alignments, single-accession-per-species alignments and tree files, bgc input files, Stacks output files and the Detarioideae bait kit sequence file are found on Dryad (<https://doi.org/10.5061/dryad.k3j9kd53w>). Data are under embargo until publication, and any further data required are available from the corresponding author upon reasonable request.

## ORCID

Rowan J. Schley  <https://orcid.org/0000-0003-1532-5353>  
 R. Toby Pennington  <https://orcid.org/0000-0002-8196-288X>  
 Oscar Alejandro Pérez-Escobar  <https://orcid.org/0000-0001-9166-2410>  
 Andrew J. Helmstetter  <https://orcid.org/0000-0003-3761-4981>  
 Manuel de la Estrella  <https://orcid.org/0000-0002-4484-3566>  
 Isabel Larridon  <https://orcid.org/0000-0003-0285-722X>  
 Timothy G. Barraclough  <https://orcid.org/0000-0002-8084-2640>  
 Félix Forest  <https://orcid.org/0000-0002-2004-433X>  
 Bente Klitgård  <https://orcid.org/0000-0002-8509-0556>

## REFERENCES

- Abbott, R. J. (2017). Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. *Journal of Systematics and Evolution*, 55, 238–258. <https://doi.org/10.1111/jse.12267>
- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26, 229–246. <https://doi.org/10.1111/j.1420-9101.2012.02599.x>
- Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, 160, 1217–1229.
- Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. Retrieved from, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Antonelli, A., & Sanmartín, I. (2011). Why are there so many plant species in the Neotropics? *Taxon*, 60(2), 403–414. <https://doi.org/10.1002/tax.602010>
- Ashton, P. S. (1969). Speciation among tropical forest trees: Some deductions in the light of recent evidence. *Biological Journal of the Linnean Society*, 1, 155–196. <https://doi.org/10.1111/j.1095-8312.1969.tb01818.x>
- Atchison, E. (1951). Studies in the Leguminosae. VI. Chromosome numbers among tropical woody species. *American Journal of Botany*, 38, 538–546.
- Backer, C. A. (1911). *Brownea hybrida* Hort. ex Backer. *Schoolflora Voor Java*, 418, 418.
- Balao, F., Casimiro-Soriguer, R., García-Castaño, J. L., Terrab, A., & Talavera, S. (2015). Big thistle eats the little thistle: Does unidirectional introgressive hybridization endanger the conservation of *Onopordum hinojense*? *New Phytologist*, 206, 448–458.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., ... Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, 19, 455–477. <https://doi.org/10.1089/cmb.2012.0021>
- Barraclough, T. G. (2019). *The Evolutionary Biology of Species*. OxfordOxford University Press.
- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59, 65–98. <https://doi.org/10.1137/141000671>
- Boecklen, W. J. (2017). Topology of Syngameons. *Ecology and Evolution*, 7, 10486–10491. <https://doi.org/10.1002/ece3.3507>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, 4, e1660. <https://doi.org/10.7717/peerj.1660>

- Buggs, R. J., & Pannell, J. R. (2006). Rapid displacement of a monoecious plant lineage is due to pollen swamping by a dioecious relative. *Current Biology*, 16, 996–1000. <https://doi.org/10.1016/j.cub.2006.03.093>
- Cannon, C. H., & Lerdau, M. T. (2015). Variable mating behaviours and the maintenance of tropical biodiversity. *Frontiers in Genetics*, 6, 183.
- Cannon, C. H., & Lerdau, M. T. (2019). Demography and destiny: The syngameon in hyperdiverse systems. *Proceedings of the National Academy of Sciences of the United States of America*, 116, 8105. <https://doi.org/10.1073/pnas.1902040116>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25, 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Cardoso, D., Särkinen, T., Alexander, S., Amorim, A. M., Bittrich, V., Celis, M., ... Forzza, R. C. (2017). Amazon plant diversity revealed by a taxonomically verified species list. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 10695–10700. <https://doi.org/10.1073/pnas.1706756114>
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, 1, 171–182.
- Crawford, T., & Nelson, E. C. (1979). Irish Horticulturists. I: *WH Crawford. Garden History*, 7, 23–26. <https://doi.org/10.2307/1586604>
- Cronk, Q. C., & Suarez-Gonzalez, A. (2018). The role of interspecific hybridization in adaptive potential at range margins. *Molecular Ecology*, 27, 4653–4656. <https://doi.org/10.1111/mec.14927>
- De La Torre, A., Ingvarsson, P. K., & Aitken, S. N. (2015). Genetic architecture and genomic patterns of gene flow between hybridizing species of *Picea*. *Heredity*, 115, 153–164. <https://doi.org/10.1038/hdy.2015.19>
- Dexter, K. G., Lavin, M., Torke, B. M., Twyford, A. D., Kursar, T. A., Coley, P. D., ... Pennington, R. T. (2017). Dispersal assembly of rain forest tree communities across the Amazon basin. *Proceedings of the National Academy of Sciences of the United States of America*, 114, 2645–2650. <https://doi.org/10.1073/pnas.1613655114>
- Dexter, K. G., Pennington, T. D., & Cunningham, C. W. (2010). Using DNA to assess errors in tropical tree identifications: How often are ecologists wrong and when does it matter? *Ecological Monographs*, 80, 267–286. <https://doi.org/10.1890/09-0267.1>
- Doyle, J. J., & Doyle, J. L. (1987). Genomic plant DNA preparation from fresh tissue- CTAB method. *Phytochemical Bulletin*, 19, 11–15.
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, 28, 2239–2252. <https://doi.org/10.1093/molbev/msr048>
- Ehrendorfer, F. (1970). Evolutionary patterns and strategies in seed plants. *Taxon*, 19, 185–195. <https://doi.org/10.2307/1217953>
- Gentry, A. H. (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. <https://doi.org/10.2307/2399084>
- Geweke, J. (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* Federal Reserve Bank of Minneapolis. Minneapolis, MN, USA Research Department.
- Gompert, Z., & Buerkle, C. A. (2010). introgress: A software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, 10, 378–384. <https://doi.org/10.1111/j.1755-0998.2009.02733.x>
- Gompert, Z., & Buerkle, C. A. (2011). Bayesian estimation of genomic clines. *Molecular Ecology*, 20, 2111–2127. <https://doi.org/10.1111/j.1365-294X.2011.05074.x>
- Gompert, Z., Lucas, L. K., Nice, C. C., Fordyce, J. A., Forister, M. L., & Buerkle, C. A. (2012). Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, 66, 2167–2181. <https://doi.org/10.1111/j.1558-5646.2012.01587.x>
- Gompert, Z., Parchman, T. L., & Buerkle, C. A. (2012). Genomics of isolation in hybrids. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, 367, 439–450. <https://doi.org/10.1098/rstb.2011.0196>
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... Paabo, S. (2010). A draft sequence of the Neandertal genome. *Science*, 328, 710–722. <https://doi.org/10.1126/science.1188021>
- Hamilton, J. A., & Aitken, S. N. (2013). Genetic and morphological structure of a spruce hybrid (*Picea sitchensis* × *P. glauca*) zone along a climatic gradient. *American Journal of Botany*, 100, 1651–1662.
- Hegarty, M. J., & Hiscock, S. J. (2009). The complex nature of allopolyploid plant genomes. *Heredity*, 103, 100–101. <https://doi.org/10.1038/hdy.2009.61>
- Ho, L. S. T., & Ané, C. (2014). A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Systematic Biology*, 63, 397–408. <https://doi.org/10.1093/sysbio/syu005>
- Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., ... Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, 330, 927–931. <https://doi.org/10.1126/science.1194585>
- Hopkins, R. (2013). Reinforcement in plants. *New Phytologist*, 197, 1095–1103. <https://doi.org/10.1111/nph.12119>
- Huson, D. H., & Bryant, D. (2005). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267. <https://doi.org/10.1093/molbev/msj030>
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., ... Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, 4, 1600016. <https://doi.org/10.3732/apps.1600016>
- Kamiya, K., Gan, Y. Y., Lum, S. K., Khoo, M. S., Chua, S. C., & Faizu, N. N. (2011). Morphological and molecular evidence of natural hybridization in *Shorea* (Dipterocarpaceae). *Tree Genetics & Genomes*, 7, 297–306. <https://doi.org/10.1007/s11295-010-0332-8>
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kearns, A. M., Restani, M., Szabo, I., Schröder-Nielsen, A., Kim, J. A., Richardson, H. M., ... Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9, 906. <https://doi.org/10.1038/s41467-018-03294-w>
- Kenzo, T., Kamiya, K., Ngo, K. M., Faizu, N., Lum, S. K. Y., Igarashi, S., ... Ichie, T. (2019). Overlapping flowering periods among *Shorea* species and high growth performance of hybrid seedlings promote hybridization and introgression in a tropical rainforest of Singapore. *Forest Ecology and Management*, 435, 38–44. <https://doi.org/10.1016/j.foreco.2018.12.038>
- Klitgaard, B. B. (1991). Ecuadorian *Brownea* and *Browneopsis* (Leguminosae-Caesalpinioideae): Taxonomy, palynology, and morphology. *Nordic Journal of Botany*, 11, 433–449. <https://doi.org/10.1111/j.1756-1051.1991.tb01244.x>
- Koptur, S. (1984). Outcrossing and pollinator limitation of fruit set: Breeding systems of neotropical *Inga* trees (Fabaceae: Mimosoideae). *Evolution*, 38, 1130–1143.
- Lawson, D. J., van Dorp, L., & Falush, D. (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications*, 9, 3258. <https://doi.org/10.1038/s41467-018-05257-7>
- Legume Phylogeny Working Group (2017). A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: *The Legume Phylogeny Working Group (LPWG)*. *Taxon*, 66, 44–77. <https://doi.org/10.12705/661.3>
- Lo, E. (2010). Testing hybridization hypotheses and evaluating the evolutionary potential of hybrids in mangrove plant

- species. *Journal of Evolutionary Biology*, 23, 2249–2261. <https://doi.org/10.1111/j.1420-9101.2010.02087.x>
- Malinsky, M. (2019). Dsuite-fast D-statistics and related admixture evidence from VCF files. *BioRxiv*, 634477.
- Mallet, J. (2007). Hybrid speciation. *Nature*, 446, 279–283. <https://doi.org/10.1038/nature05706>
- Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, 34, 531–544. <https://doi.org/10.1016/j.tree.2019.02.008>
- Mayr, E. (1942). *Systematics and the Origin of Species*. New York, New York, USA: Columbia University Press.
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, 8, 14363. <https://doi.org/10.1038/ncomms14363>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *Proceedings of the Gateway Environments Workshop (GCE)*, 1–8.
- Misiewicz, T. M., & Fine, P. V. (2014). Evidence for ecological divergence across a mosaic of soil types in an Amazonian tropical tree: *Protium subseratum* (Burseraceae). *Molecular Ecology*, 23, 2543–2558.
- Morales-Briones, D. F., Liston, A., & Tank, D. C. (2018). Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytologist*, 218, 1668–1684.
- Nevado, B., Contreras-Ortiz, N., Hughes, C., & Filatov, D. A. (2018). Pleistocene glacial cycles drive isolation, gene flow and speciation in the high-elevation Andes. *New Phytologist*, 219, 779–793. <https://doi.org/10.1111/nph.15243>
- Ojeda, D. I., Koenen, E., Cervantes, S., de la Estrella, M., Banguera-Hinestroza, E., Janssens, S. B., ... Hardy, O. J. (2019). Phylogenomic analyses reveal an exceptionally high number of evolutionary shifts in a florally diverse clade of African legumes. *Molecular Phylogenetics and Evolution*, 137, 156–167. <https://doi.org/10.1016/j.ympev.2019.05.002>
- Page, A. J., Taylor, B., Delaney, A. J., Soares, J., Seemann, T., Keane, J. A., & Harris, S. R. (2016). SNP-sites: Rapid efficient extraction of SNPs from multi-FASTA alignments. *Microbial Genomics*, 2. <https://doi.org/10.1099/mgen.0.000056>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for stacks. *Methods in Ecology and Evolution*, 8, 1360–1373.
- Payseur, B. A. (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, 10, 806–820. <https://doi.org/10.1111/j.1755-0998.2010.02883.x>
- Pennington, R. T., & Lavin, M. (2016). The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytologist*, 210, 25–37. <https://doi.org/10.1111/nph.13724>
- Pérez, Á. J., Klitgård, B. B., Saslis-Lagoudakis, C., & Valencia, R. (2013). *Brownea jaramilloi* (Leguminosae: Caesalpinioideae), a new, overlooked species endemic to the Ecuadorian Amazon. *Kew Bulletin*, 68, 157–162. <https://doi.org/10.1007/s12225-012-9423-z>
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One*, 7, e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Petit, R. J., Bodénès, C., Ducouso, A., Roussel, G., & Kremer, A. (2004). Hybridization as a mechanism of invasion in oaks. *New Phytologist*, 161, 151–164. <https://doi.org/10.1046/j.1469-8137.2003.00944.x>
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC. *R News*, 6, 7–11.
- Python Software Foundation. (2010) Python Language Reference, version 2.7. Retrieved from <http://www.python.org>
- R Development Core Team. (2013). *R: A language and environment for statistical computing*. Retrieved from <http://www.R-project.org/>
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197, 573–589. <https://doi.org/10.1534/genetics.114.164350>
- Rambaut, A., Suchard, M. A., Xie, D., & Drummond, A. J. (2015) Tracer. Retrieved from <http://tree.bio.ed.ac.uk/software/tracer/>, 1.6.
- Rieseberg, L. H., Kim, S., Randell, R. A., Whitney, K. D., Gross, B. L., Lexer, C., & Clay, K. (2007). Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica*, 129, 149–165. <https://doi.org/10.1007/s10709-006-9011-y>
- Rieseberg, L. H., & Wendel, J. F. (1993). Introgression and its consequences in plants. *Hybrid Zones and the Evolutionary Process*, 70, 70–109.
- Royer, A. M., Streisfeld, M. A., & Smith, C. I. (2017). Data from: Population genomics of divergence within an obligate pollination mutualism: Selection maintains differences between Joshua tree species. *Dryad*. <https://doi.org/10.5061/dryad.7pj4t>
- Schley, R. J., de la Estrella, M., Pérez-Escobar, O. A., Bruneau, A., Barraclough, T., Forest, F., & Klitgård, B. (2018). Is Amazonia a ‘museum’ for Neotropical trees? The evolution of the *Brownea* clade (Detarioideae, Leguminosae). *Molecular Phylogenetics and Evolution*, 126, 279–292. <https://doi.org/10.1016/j.ympev.2018.04.029>
- Scotti-Saintagne, C., Dick, C. W., Caron, H., Vendramin, G. G., Guichoux, E., Buonamici, A., ... Lemes, M. R. (2013). Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae). *Journal of Biogeography*, 40, 676–692.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15, 176–192. <https://doi.org/10.1038/nrg3644>
- Smith, S. A., Moore, M. J., Brown, J. W., & Yang, Y. (2015) Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, 15. <https://doi.org/10.1186/s12862-0>
- Solís-Lemus, C., & Ané, C. (2016). Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics*, 12, e1005896. <https://doi.org/10.1371/journal.pgen.1005896>
- Solís-Lemus, C., Bastide, P., & Ané, C. (2017). PhyloNetworks: A package for phylogenetic networks. *Molecular Biology and Evolution*, 34, 3292–3298. <https://doi.org/10.1093/molbev/msx235>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Stenz, N. W., Larget, B., Baum, D. A., & Ané, C. (2015). Exploring tree-like and non-tree-like patterns using genome sequences: An example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. *Systematic Biology*, 64, 809–823.
- Stift, M., Kolář, F., & Meirmans, P. G. (2019). STRUCTURE is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity*, 123, 429–441. <https://doi.org/10.1038/s41437-019-0247-6>
- Struck, T. H., Feder, J. L., Bendiksbj, M., Birkeland, S., Cerca, J., Gusarov, V. I., ... Dimitrov, D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, 33, 153–163. <https://doi.org/10.1016/j.tree.2017.11.007>
- Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: A plant perspective. *Biology Letters*, 14. <https://doi.org/10.1098/rsbl.2017.0688>

- Sullivan, A. R., Owusu, S. A., Weber, J. A., Hipp, A. L., & Gailing, O. (2016). Hybridization and divergence in multi-species oak (*Quercus*) communities. *Botanical Journal of the Linnean Society*, 181, 99–114.
- Sweigart, A. L., & Willis, J. H. (2003). Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution*, 57, 2490–2506. <https://doi.org/10.1111/j.0014-3820.2003.tb01494.x>
- Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, 3, 170–177. <https://doi.org/10.1038/s41559-018-0777-y>
- ter Steege, H., Pitman, N. C. A., Sabatier, D., Baraloto, C., Salomao, R. P., Guevara, J. E., ... Silman, M. R. (2013). Hyperdominance in the Amazonian tree flora. *Science*, 342, 1243092. <https://doi.org/10.1126/science.1243092>
- Ulloa Ulloa, C., Acevedo-Rodríguez, P., Beck, S., Belgrano, M. J., Bernal, R., Berry, P. E., ... Jørgensen, P. M. (2017). An integrated assessment of the vascular plant species of the Americas. *Science*, 358, 1614–1617. <https://doi.org/10.1126/science.aao0398>
- Valencia, R., Balslev, H., & Miño, G. P. Y. (1994). High tree alpha-diversity in Amazonian Ecuador. *Biodiversity & Conservation*, 3, 21–28. <https://doi.org/10.1007/BF00115330>
- Vonlanthen, P., Bittner, D., Hudson, A. G., Young, K. A., Müller, R., Lundsgaard-Hansen, B., ... Seehausen, O. (2012). Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature*, 482, 357–362. <https://doi.org/10.1038/nature10824>
- Whitney, K. D., Randell, R. A., & Rieseberg, L. H. (2010). Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, 187, 230–239. <https://doi.org/10.1111/j.1469-8137.2010.03234.x>
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis* Springer.
- Yang, W., While, G. M., Laakkonen, H., Sacchi, R., Zuffi, M. A. L., Scali, S., ... Uller, T. (2018). Genomic evidence for asymmetric introgression by sexual selection in the common wall lizard. *Molecular Ecology*, 27, 4213–4224. <https://doi.org/10.1111/mec.14861>
- Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, 19, 153. <https://doi.org/10.1186/s12859-018-2129-y>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Schley RJ, Pennington RT, Pérez-Escobar OA, et al. Introgression across evolutionary scales suggests reticulation contributes to Amazonian tree diversity. *Mol Ecol*. 2020;00:1–16. <https://doi.org/10.1111/mec.15616>