Evolutionary reticulation makes an important contribution to Amazonian tree diversity

Rowan J. Schley[1,2*], R. Toby Pennington[3,4], Oscar Alejandro Pérez-Escobar[1], Andrew J. Helmstetter[5], Manuel de la Estrella[6], Isabel Larridon[1,7], Izai Alberto Bruno Sabino Kikuchi[1,8], Timothy Barraclough[2], Félix Forest[1], and Bente Klitgård[1]

[1]*Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK*

[2]*Department of Life Sciences, Imperial College London, Silwood Park, Ascot, Berkshire, UK, SL5 7PY*

[3] *Geography, University of Exeter, Laver Building, North Park Road, Exeter EX4 4QE*

[4]*Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh EH3 5LR*

[5]*Institut de Recherche pour le Développement (IRD), UMR-DIADE, BP 64501, F-34394 Montpellier cedex 5, France.*

[6]*Departamento de Botánica, Ecología y Fisiología Vegetal, Facultad de Ciencias, Campus de Rabanales, Universidad de Córdoba, 14071, Córdoba, Spain*

[7] *Systematic and Evolutionary Botany Lab, Department of Biology, Ghent University, K.L. Ledeganckstraat 35, 9000 Gent, Belgium*

[8]*Universiteit Leiden, Hortus botanicus Leiden, PO Box 9500, Leiden, 2300 RA, The Netherlands*

*Corresponding author at: *Room E.2.6, Herbarium, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3AE, UK*

E-mail address: r.schley@kew.org *(R. J. Schley).*

Abstract

Hybridization has the potential to generate or homogenize biodiversity and is a particularly common phenomenon in plants, with an estimated 25% of species undergoing inter-specific gene flow. However, hybridization has rarely been demonstrated among tree species in Amazonia, the world's largest rainforest, and its role in diversification remains enigmatic. We show that within *Brownea*, a characteristic tree genus of Amazonia, there is extensive evidence of hybridization. Using both phylogenomic and population genomic approaches we find multiple historical hybridization events within *Brownea*, along with contemporary hybridization among co-occurring species. Finally, we infer homogeneous rates of gene flow among different genomic loci, indicating a lack of selection against hybrids, reflecting their persistence over time. These results demonstrate that gene flow between Amazonian tree species has occurred across temporal scales, which contrasts with previous

work on Amazonian trees, and may have contributed to the evolution of the most diverse tree flora on Earth.

Main

Reproductive isolation is often seen as a prerequisite for speciation and as a defining feature of species [1, 2]. Despite this, hybridization between species is known to occur, and has several different outcomes, from the erosion of evolutionary divergence [3, 4] to the formation of entirely new 'hybrid species' [5]. Neotropical rainforests harbour the highest levels of plant diversity on Earth [6], and 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 [7-9]. Although a few observations of reproductive biology support this, such as the intersterility between lineages of *Inga*, a species-rich genus in the legume family [10], evidence for hybridization has been poorly tested empirically in tropical floras. The only potential example of hybridization among neotropical trees which has been documented using DNA sequence data involves two species of *Carapa* [11] found in Amazonia, the largest expanse of rainforest in the world which contains at least 6,800 tree species [12].

One factor that suggests hybridization might occur more frequently in neotropical rainforest tree species, if their reproductive isolation is not absolute, is the remarkable level of sympatry found for closely related species. In several neotropical lineages, including *Inga*, *Guatteria* and *Swartzia*, many recently diverged species co-occur [13], 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 [14]. 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, allowing genetic isolation between species to be maintained (i.e., reinforcement [15]). In other cases, there may be a permanent movement of genetic material from one lineage to another [16], which is known as

‘introgression’. This transfer of genetic material through hybridization may confer a selective advantage to resultant offspring [17], in which case it is referred to as ‘adaptive introgression’. Adaptive introgression is most often observed between closely-related taxa during the invasion of new habitats [18, 19]. 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 [20]. This has been demonstrated in a wide range of taxa, including sunflowers, African cichlid fish and Darwin’s finches [21-23].

Introgression can occur at different rates in different regions of the genome [24]. Regions under divergent selection may remain distinct due to reduced fitness of hybrid genotypes at such loci, resulting in a low rate of introgression. Conversely, regions under little or no selection tend to introgress more freely, becoming homogenised 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 [25]. Such a process has been demonstrated in temperate-zone tree species, where divergence between hybridizing lineages is maintained through environmentally-driven selection [26, 27]. This explains why species that hybridize can remain as biologically distinct (and taxonomically identifiable) groups despite undergoing genetic exchange with other species [28, 29].

Brownea, a member of the pea and bean family (Fabaceae/Leguminosae), is a characteristic tree genus of lowland neotropical rainforests, and contains around 27 species distributed across northern South America. Previous work indicates that there is a broad degree of phylogenetic incongruence evident in this genus [30] which might indicate hybridization, although this could also result from incomplete lineage sorting (the differential inheritance of genetic variation from a common ancestral species in different descendent lineages). However, there are numerous *Brownea* hybrids in cultivation (e.g. *B. x crawfordii* [31] and *B. hybrida* [32]), 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* [33] and the wide-ranging *B. grandiceps*, which co-occur in the Ecuadorian Amazon (Fig. 1). There are multiple morphological distinctions between these two species, including differences in inflorescence colour and structure, growth habit, tree height and leaf morphology [33]. 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 [33, 34]. Despite this, hybridization appears to occur, as evidenced by the existence of a putative hybrid between these two species known as *B.* "rosada" (Fig. 1). *Brownea "*rosada*"* displays an intermediate morphology between its two parental species, producing pink flowers. The hypothesis of a *B. jaramilloi* x *B. grandiceps* hybrid has not yet been tested, however, using molecular data.

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).

As a member of the legume family, which dominates Amazonian forests [35], 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 neotropical rainforest trees. Systematically documenting hybridization at a range of time scales and taxonomic levels within this group could reveal how admixture has contributed to the assembly of one of the world's richest floras. Accordingly, this investigation aims to answer the following questions:

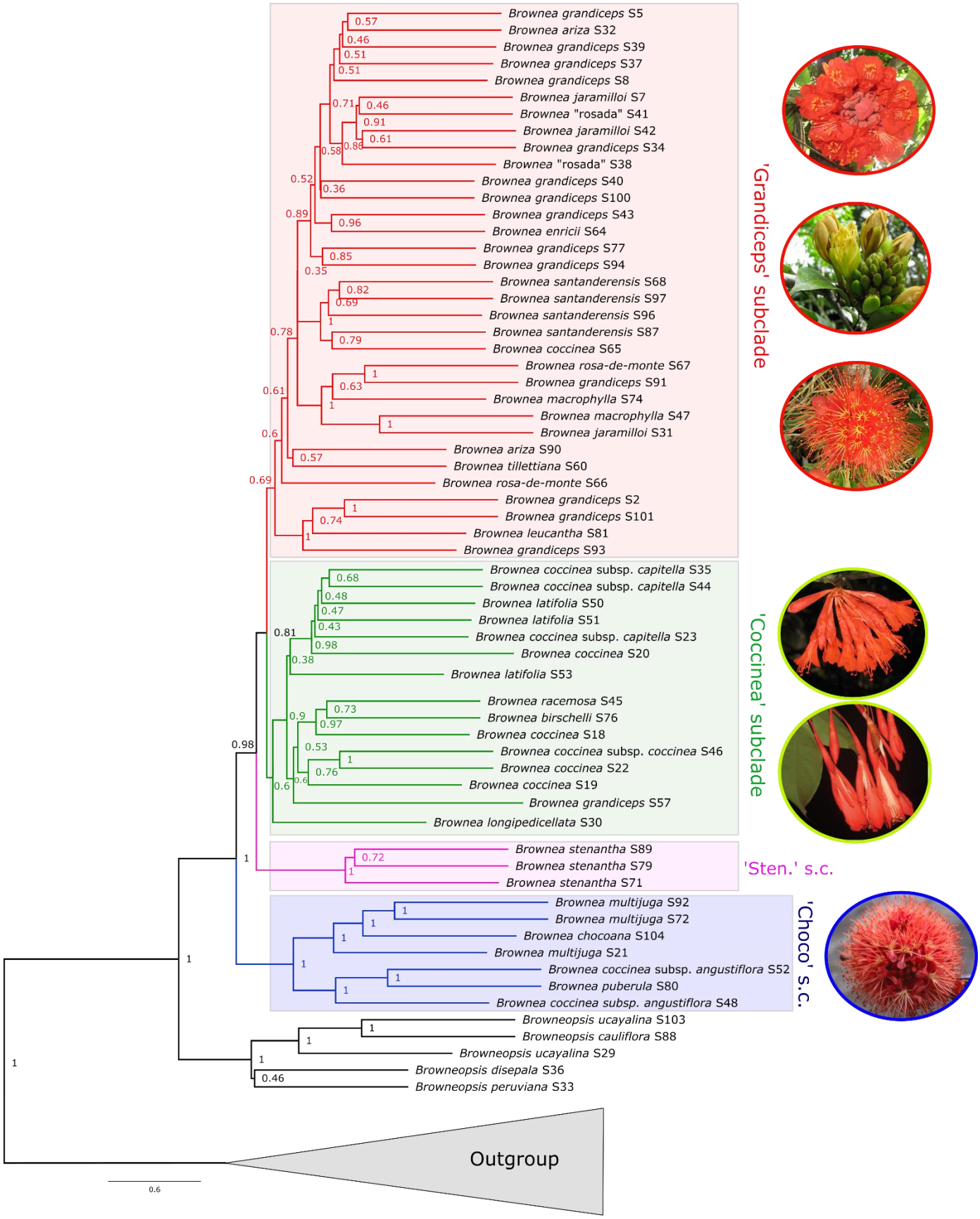
1. Is there evidence of hybridization at a deep phylogenetic level in this lineage of neotropical rainforest tree?
2. Is there evidence of more recent gene flow, and if so, does this occur evenly across most of the genome?

Results

Phylogenetic analysis

Using DNA sequence data from 226 nuclear gene regions, we produced a 'species' tree using *ASTRAL* [36] (Fig. 2) based on gene trees inferred with the program *RAxML* [37]. This resulted in well-supported relationships between major subclades (>0.9 local posterior probability (PP)), with lower support for inter-specific relationships. *ASTRAL* 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 Supplementary Fig. 1.

Figure 2: *ASTRAL* species tree inferred from *RAxML* gene trees using the multi-species coalescent model. Numbers at nodes of the tree signify local posterior probability (LPP) values, a measure of support for each quadripartition. Taxa within the red box belong to the 'Grandiceps' subclade, taxa within the green box belong to the 'Coccinea' subclade, taxa within the pink box represent the 'Stenantha' subclade and those within the blue box belong to the 'Chocóan' subclade. 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).


Brownea grandiceps S5
Brownea ariza S32
Brownea grandiceps S39
Brownea grandiceps S37
Brownea grandiceps S8
Brownea jaramilloi S7
Brownea "rosada" S41
Brownea jaramilloi S42
Brownea grandiceps S34
Brownea "rosada" S38
Brownea grandiceps S40
Brownea grandiceps S100
Brownea grandiceps S43
Brownea enricii S64
Brownea grandiceps S77
Brownea grandiceps S94
Brownea santanderensis S68
Brownea santanderensis S97
Brownea santanderensis S96
Brownea santanderensis S87
Brownea coccinea S65
Brownea rosa-de-monte S67
Brownea grandiceps S91
Brownea macrophylla S74
Brownea macrophylla S47
Brownea jaramilloi S31
Brownea ariza S90
Brownea tillettiana S60
Brownea rosa-de-monte S66
Brownea grandiceps S2
Brownea grandiceps S101
Brownea leucantha S81
Brownea grandiceps S93
'Grandiceps' subclade
Brownea coccinea subsp. capitella S35
Brownea coccinea subsp. capitella S44
Brownea latifolia S50
Brownea latifolia S51
Brownea coccinea subsp. capitella S23
Brownea coccinea S20
Brownea latifolia S53
Brownea racemosa S45
Brownea birschelli S76
Brownea coccinea S18
Brownea coccinea subsp. coccinea S46
Brownea coccinea S22
Brownea coccinea S19
Brownea grandiceps S57
Brownea longipedicellata S30
'Coccinea' subclade
Brownea stenantha S89
Brownea stenantha S79
Brownea stenantha S71
'Sten.' s.c.
Brownea multijuga S92
Brownea multijuga S72
Brownea chocoana S104
Brownea multijuga S21
Brownea coccinea subsp. angustiflora S52
Brownea puberula S80
Brownea coccinea subsp. angustiflora S48
'Choco' s.c.
Browneopsis ucayalina S103
Browneopsis cauliflora S88
Browneopsis ucayalina S29
Browneopsis disepala S36
Browneopsis peruviana S33
Outgroup
0.6

130

131

Ancient introgression in Brownea species

Phylogenetic networks were estimated using concordance factors (CF) generated from *RAxML* gene trees to ascertain whether the historical pattern of diversification in *Brownea* was tree-like or reticulate. Networks with different numbers of hybridization events (*h*) were compared using -log pseudolikelihoods. These comparisons pinpointed two hybridization events within *Brownea* (i.e., the best-fit network model was $h = 2$, as indicated by the -log pseudolikelihood values in Supplementary Fig. 2). Negative log pseudolikelihoods increased steadily between $h = 0$ and $h = 2$, after which the increasing number of hybridization events only make minimal improvements to -log pseudolikelihood.

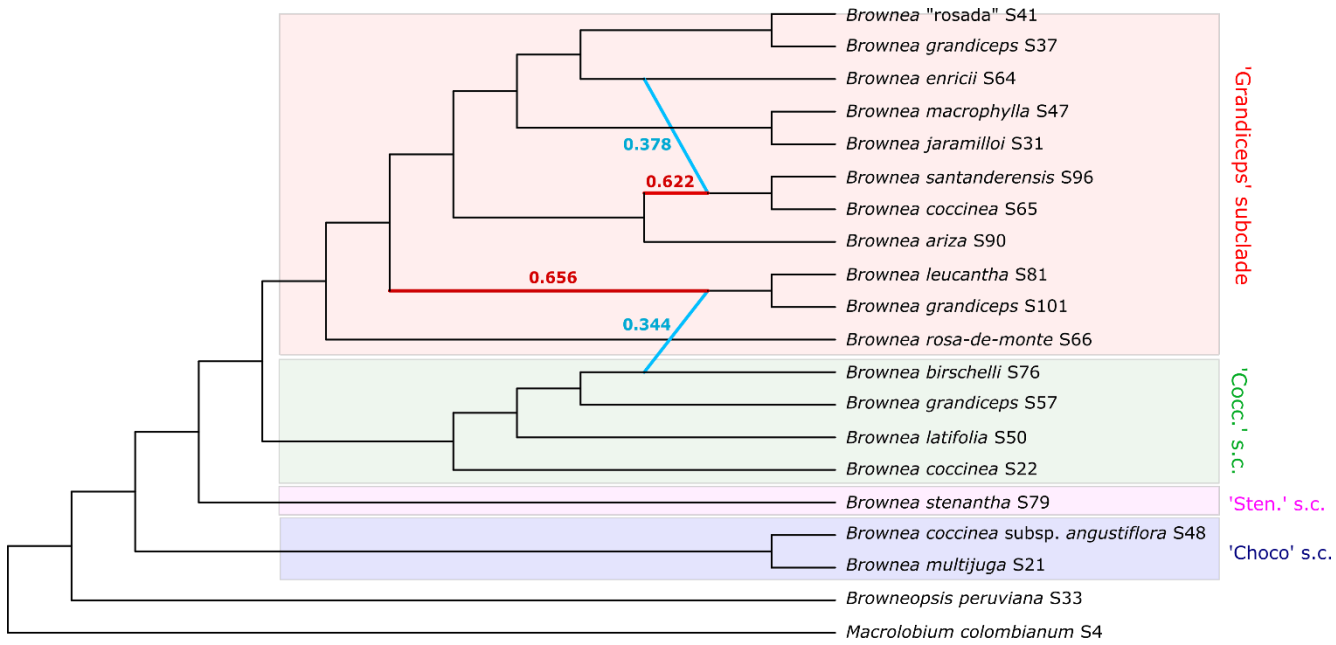


Figure 3: Phylogenetic network with two hybridization events (h = 2), estimated using *SNaQ!* in the Julia package *PhyloNetworks*. Light blue horizontal branches indicate inferred hybridization events, and numbers next to the branches show the estimated proportion of genes contributed by each lineage 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. Coloured boxes containing taxa are described in the legend of Fig. 2.

The inferred phylogenetic network (Fig. 3) showed a broadly similar topology to the species tree in Fig. 2, with the addition of two hybridization events. 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 occurs 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.

TICR (Tree Incongruence Checking in R) analysis indicated that a network-like model 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 x 10^{-19}, X^2 = 91.149), which differed significantly from the CF values expected under a model containing 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.

Introgression dynamics across loci

Having examined the degree of historical reticulation among *Brownea* lineages, 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 visualised for co-occurring individuals of *B. grandiceps* and *B. jaramilloi*, which putatively hybridize in the wild. Following this, in order 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.

In total, ddRAD sequencing resulted in 640,898,950 reads between 350-550bp 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*) (Supplementary Table 3). The 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 were visualised using principal component analysis (PCA) and a genetic distance network inferred with the program *SPLITSTREE* [38]. These analyses indicated a distinct signature of admixture between *B. grandiceps* and *B. jaramilloi*. The PCA of SNP variation inferred using the R package *ggplot2* (Fig. 4) 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* (Supplementary Fig. 3). *SPLITSTREE* recovered a clustering of individuals into two main groups, largely representing *B. grandiceps* and *B. jaramilloi*, with four putative hybrid individuals displaying an intermediate relationship between the two species clusters.

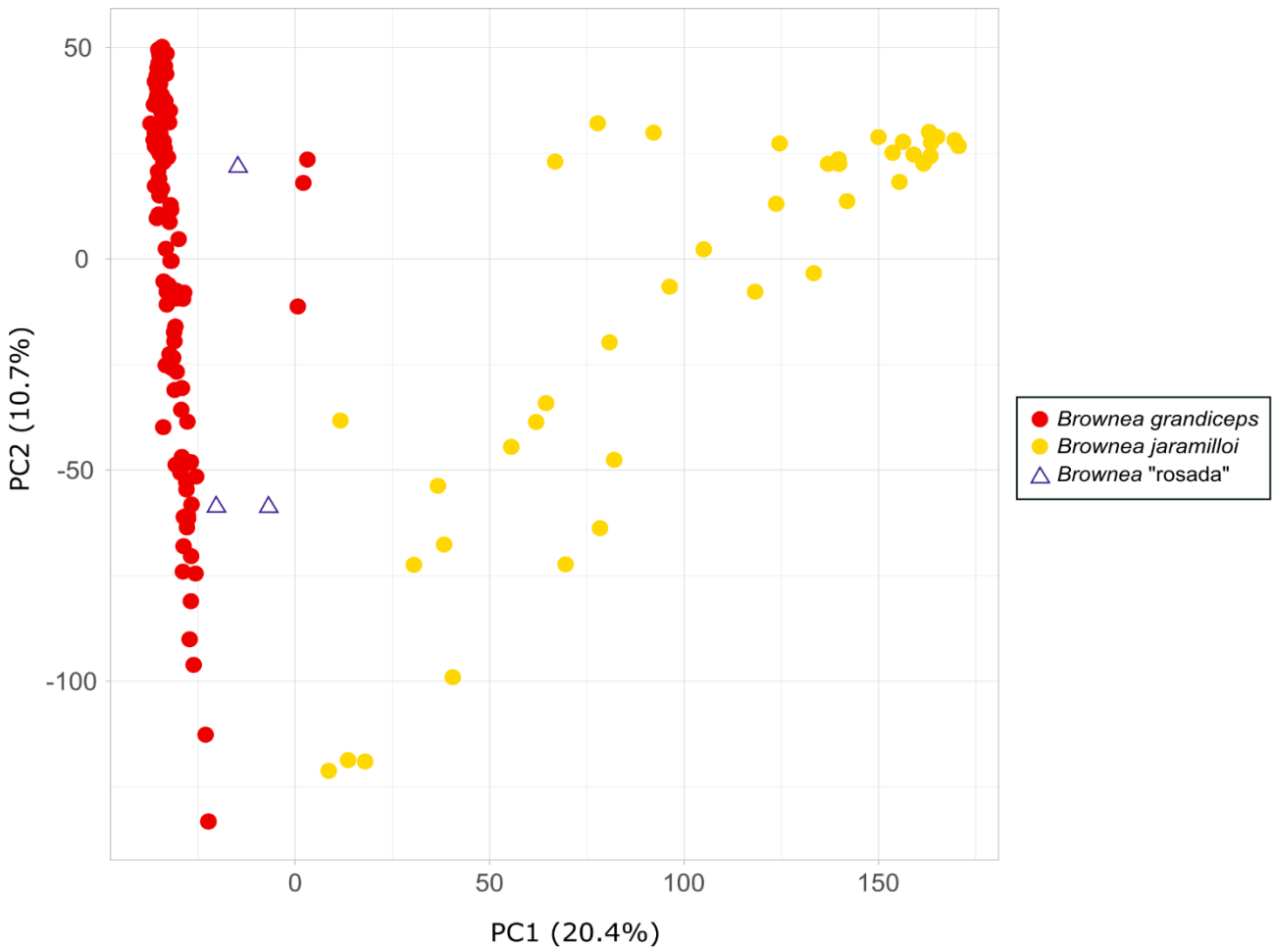


Figure 4: Principal component analysis of genotype data for all SNPs inferred at all loci. Individuals are coded by colour and point shape: red circles denote individuals identified as *B. grandiceps*, yellow circles denote individuals identified as *B. jaramilloi*, and blue triangles denote putative hybrid individuals (*B.* "rosada"). The amount of the genetic variation explained by each principle component is shown next to axes, which are labelled PC1 and PC2.

A *fastSTRUCTURE* analysis [39] was performed to further examine the degree of shared genetic variation, as well as providing information on the population genetic structure of the two species. This analysis indicated that there was a large amount of shared ancestry between the genotyped individuals (Fig. 5), 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 populations) was two ($\Delta_{\text{marginal likelihood}} = 0.085$, Supplementary Fig. 4A), which was the value that resulted in the largest increase in marginal likelihood. Since the 'best' value of *K* is an estimate, f*astSTRUCTURE* plots generated with other *K* values are shown in Supplementary Fig. 4B. Figure 5 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 (Supplementary Fig. 4C).

Figure 5: *fastSTRUCTURE* plot, indicating ancestry proportions for two populations (*K* = 2). The 'populations' correspond to two different species: *B. grandiceps* (shown in red) and *B. jaramilloi* (shown in yellow). Each individual accession is represented by a column, and the proportion of ancestry from either species (mean *Q*) is proportional to the length of different coloured bars in each column. Individuals identified as *B.* "rosada" are marked with a blue box and a blue triangle beneath their column. All taxa to the left and right of these boxes were identified as *B. jaramilloi* and *B. grandiceps*, respectively.

A *NEWHYBRIDS* analysis [40] (https://github.com/eriqande/newhybrids) was used to categorize 169 of 171 genotyped individuals into different hybrid classes (pure, F1, F2, and backcrosses). This was performed on three subsets of 500 loci due to the computational limitations of the program. *NEWHYBRIDS* 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%. 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%). In the latter, the probability of any one hybrid class did not exceed 90%, which was the threshold used to categorize individuals as belonging to a certain class.

Bayesian estimation of genomic clines (*bgc*) was utilised to determine whether any loci showed outlying rates of admixture relative to the genome-wide average among all genotyped individuals. This method estimates two parameters for each locus: α (the 'direction' of introgression) and β (the 'rate' of introgression), which were used to infer the impact of selection on introgressing loci. *bgc* 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 mostly *B. jaramilloi* alleles among all individuals (i.e. with positive α estimates) and 1089 loci (5.69%) with mostly *B. grandiceps* alleles among all individuals (i.e., with negative α estimates). However, no loci displayed extreme rates of introgression relative to the average rate across the genome (i.e., there were no statistically outlying β parameter estimates). The MCMC runs from which these results were drawn showed adequate mixing (Supplementary Fig. 5A-C).

Discussion

Our study represents the first clear case of reticulated 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. We demonstrate that within *Brownea*, a

characteristic Amazonian tree genus, reticulated evolution has occurred over the course of its evolutionary history, with evidence of introgression deep in the history of the genus and more recently, between the closely related *B. grandiceps* and *B. jaramilloi*.

Phylogenetic network analysis suggested that introgression has taken place between *Brownea* lineages in the past, with two separate hybridization events inferred and displayed in Fig. 3. The concordance factors obtained from gene trees were not adequately explained by a tree-like model (i.e., one accounting only for incomplete sorting of ancestral variation), suggesting that a network-like model (i.e., one including introgression) best describes the diversification patterns within *Brownea* (P= 1.240 x 10^{-19}, X^2 = 91.149). Hybridization mainly occurs between individuals which are geographically overlapping, or have been so at some point in the past [41]. Our findings are thus further corroborated by the fact that both historical introgression events were between taxa that co-occur (e.g., *B. grandiceps, B. leucantha* and *B. birschelli* on the Northern coast of Venezuela, and *B. coccinea, B. santanderensis* and *B. enricii* within the Colombian cordilleras). This is also in agreement with previous work [30], which indicated that the 'stem' lineages within *Brownea* shared a broad distribution across northern South America, which may have further facilitated earlier hybridization. Our study also recovered a low degree of support for bifurcating inter-specific relationships (Fig. 2) and a high degree of incongruence among gene trees (Supplementary Fig. 1). These results can be partially explained by introgression, rather than being caused purely by the stochasticity of lineage sorting in large populations, which is suggested to be a common phenomenon in many rainforest trees with large population sizes[42]. It seems that amongst closely-related species of *Brownea* it is likely that divergence has not progressed to the point of complete reproductive isolation, as has been shown in temperate *Quercus* species[43]. Moreover, 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 [44].

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 in the past, before the

divergence of the descendent species. The inferred 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 [30]. Accordingly, it is unlikely that the inferred hybridization events are due to 'accidental' hybridization between co-occurring species, merely producing maladapted F1 hybrids which will eventually be outcompeted, fortifying the boundaries between species (a process known as 'reinforcement' [15]). Rather, our results suggest a persistence of introgression over evolutionary time within *Brownea*.

This investigation also uncovered a substantial signal of recent introgression between *B. grandiceps* and *B. jaramilloi* in the Yasuní National Park 50ha plot located in the Ecuadorian Amazon, with evidence of multi-generational 'backcrossing', further suggesting that hybridization does not always result in reinforcement. All hybrid individuals exhibited some degree of backcrossing in the *NEWHYBRIDS* analysis, which implies that hybrids persist over time. The low F_{st} estimate for the *B. grandiceps/B.* "rosada" and *B. jaramilloi* populations (0.11), in addition to the principal component analysis (Fig. 4), the *fastSTRUCTURE* analysis (Fig. 5) and the *SPLITSTREE* plot (Supplementary Fig. 3) 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 identified as *B. jaramilloi*, although hybrid individuals are also observed in the *B. grandiceps* cluster. The higher number of variant sites and the higher nucleotide diversity (π) recovered for *B. jaramilloi* (Supplementary Table 3) 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 [45].

The *bgc* analysis indicated that most introgression was asymmetric, with more loci containing mostly *B. grandiceps* alleles (5.96% of all loci) than loci containing mostly *B. jaramilloi* alleles (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 β parameter. The excess of *B. grandiceps*

ancestry mirrors the asymmetry of introgression suggested by Fig. 5 and is likely driven by the uneven population sizes of the two species. *Brownea jaramilloi* has only ever been observed in a small part of the western Amazon [33] and is much less populous than *B. grandiceps*, which occurs across northern South America. More specifically, 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 [46]. Pollen swamping can serve as a mechanism of invasion (e.g., in *Quercus* [47]), and the associated asymmetric introgression may lead to the extinction of the rarer species, especially when hybrids are inviable or sterile [48]. 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. The best evidence of this is the absence of loci with extreme values of the β parameter recovered by our *bgc* analysis. Extreme deviations in β are mainly expected in the presence of gene flow when selection against hybrids is strong, resulting in underdominance and a paucity of heterozygous sites [25].

Further to this, the observed asymmetry in introgression could be caused by selection favouring hybrid genotypes. The transfer of genomic regions which confer a selective advantage from a donor to a recipient species (i.e. adaptive introgression) can result in a disproportionate amount of one species' genome being present in hybrid individuals [49]. This occurs when viable hybrids only tend to backcross with one of the parental species, thereby introducing the other species' genetic material into the first in an asymmetrical fashion. Selection has been suggested as a driver for asymmetric introgression in multiple plant genera, including *Helianthus* [50] and *Iris* [51]. Accordingly, it is possible that selectively advantageous alleles are passing between species (e.g., between *B. jaramilloi* and *B. grandiceps*) through backcrossing events over many generations as observed by the *fastSTRUCTURE* and *NEWHYBRIDS* analyses in this study [52]. This may facilitate adaptation to new niches due to the widening of the pool of variation upon which selection can act [18]. However, it is difficult to ascertain whether adaptive introgression has occurred without measuring the impact of introgression on variation in the phenotype and its fitness effects [19].

Studies such as ours that document hybridization between tree species within Amazonia, and within tropical rainforests in general, are rare. Indeed, it was previously suggested that inter-specific hybrids between rainforest tree species are poor competitors, and that fertile hybrid populations are nearly non-existent [8]. While there is some evidence of introgression in tropical trees[11], most available studies substantiating this are based on trees from other tropical regions or habitats (e.g., *Shorea* in Asia [53, 54], or *Rhizophora* in Indo-Pacific mangroves [55]). 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 found evidence of introgression across taxonomic, spatial and temporal scales, with evidence of backcrossing and a lack of selection against hybrid genotypes. 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 help explain why we found evidence of introgression both at macroevolutionary and microevolutionary scales. While it is difficult to determine whether hybridization between Amazonian tree lineages is a more common occurrence than previously thought, or whether it is a tendency unique to certain rainforest genera such as *Brownea*, it may be prudent to review how relationships among tropical tree lineages are inferred. Accordingly, in order to understand the evolutionary relationships within such groups using phylogenetic networks 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 [41, 56].

Methods

Phylogenetic analysis

Sequences from 226 nuclear genes were used to elucidate the evolutionary relationships between *Brownea* species using a phylogenomic approach. DNA sequence data were generated from leaf material collected from herbarium specimens and silica-gel dried accessions using targeted bait capture. Targeted bait capture uses RNA probes to enable the sequencing of fragmented DNA, such as

that collected from historical natural history collections. Details of library preparation, hybridization and sequencing are detailed in Supplementary Methods 1. Twenty-three of twenty-seven lineages were sampled within the genus *Brownea*, including the three subspecies of *B. coccinea* and one putative hybrid lineage (*Brownea* "rosada", a putative hybrid 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 *Macrolobium*, *Heterostemon*, *Paloue* and *Browneopsis*, which form part of the 'Brownea clade' (Leguminosae, subfamily Detarioideae [57]). The list of accessions and their associated information can be found in Supplementary Table 1.

DNA sequencing reads were quality-checked with the program *FASTQC* v0.11.3 [58], and were subsequently trimmed using *Trimmomatic* v.0.3.6 [59]. This was done to remove adapter sequences and to quality-filter reads, by removing reads that were too short or of poor quality, as assessed by their phred-33 score. Following quality-filtering, loci were assembled using *SPAdes* v3.11.1 [16] by the HybPiper pipeline v1.2 [60], and potentially paralogous loci were removed using the Python [61] script '*paralog_investigator.py*', distributed with the HybPiper pipeline. All sequences were aligned by gene region using *MAFFT* v7.215 [62]. In order to infer gene trees for phylogenetic network analysis, the 226 gene regions were further refined to include only 20 taxa, representing a single accession per lineage in *Brownea*, using *Macrolobium colombianum* as the outgroup taxon. Where applicable, samples were chosen by comparing the sequence recovery of conspecific accessions and choosing the individual with the best gene recovery. This resulted in 220 single-accession-per-lineage gene alignments. Further details of quality filtering, read assembly and alignment of DNA sequencing reads are given in Supplementary Methods 1.

Gene trees were generated for each of the 226 gene alignments using *RAxML* v.8.0.26 [37] with 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 datasets. *Macrolobium colombianum* was used to root both the full and single-accession dataset analyses since it was the *Macrolobium* accession that was present in the most gene alignments. The gene trees from both datasets were used to generate species trees under the multi-species coalescent model in *ASTRAL* v.5.6.1 [36]. The heuristic version of *ASTRAL*

was used because gene trees were estimated for >30 taxa, and the default parameters for the program were used for species tree inference. Monophyly was not enforced for individuals belonging to the same species (the '*-a*' option in *ASTRAL*). Finally, discordance between gene trees was calculated and visualised for the full dataset using the Java program *PhyParts* v0.0.1 [63] (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).

Inferring ancient introgression

Phylogenetic networks were inferred for 220 gene trees from the single-accession-per-lineage dataset to understand whether introgression occurred over the course of the evolutionary history of *Brownea*. Networks were inferred with the program *SNaQ!*, implemented in the *Julia* v0.6.4[64] package *PhyloNetworks* v0.11.0 [56]. Networks were estimated using gene trees from the single-accession-per-lineage dataset as per the assumptions of the program, which requires that each tip of the phylogenetic trees represent a single lineage. This dataset contained 20 taxa from the genera *Brownea*, *Browneopsis* and *Macrolobium*. The network with the number of hybridization events (*h*) best describing the data was chosen using negative log pseudolikelihood comparison. Finally, the topologies of the gene trees were compared with those of the best-fit phylogenetic network 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 [65], implemented in the R v3.4.4 [66] package *PHYLOLM* [67]. 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 Supplementary Methods 1.

Population-level introgression

The architecture and rates of introgression across the genome were examined using two ecologically divergent *Brownea* species which co-occur in the Ecuadorian Amazon. In order to do this we used ddRADseq [68], a reduced-representation genotyping approach which uses restriction enzymes to fragment genomic DNA to generate many thousands of sequence markers *de-novo* for SNP discovery. One-hundred and seventy-one specimens in total were genotyped using dried leaf material. 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 was collected from *Brownea* trees in the Yasuní National Park 50ha forest plot in Napo, Ecuador, and dried in a herbarium press. The sample list is shown in Supplementary Table 2. Sampling, library preparation and sequencing protocols are explained further in Supplementary Methods 2.

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 [69]. 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.
This resulted in a dataset containing 22,046 loci with 120,085 SNPs for 171 individuals. A dataset 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 dataset containing 19,130 loci with 19,130 SNPs for 171 individuals. Details of read assembly and filtering are shown in Supplementary Methods 2.

In order to understand the patterns of introgression at the population level, the full ddRAD dataset was used to visualise 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* [70], and by producing a neighbour net plot in *SPLITSTREE* v4.14.6 [38]. Furthermore, the single-SNP-per-locus dataset containing 19,130 RAD loci was used to infer population structure with the program *fastSTRUCTURE* v1.0 [39]. The number of populations (*K*) was chosen using the value which provided the largest improvement in marginal likelihood. Finally, *NEWHYBRIDS* v1.1 [40] (https://github.com/eriqande/newhybrids) was used to categorise individuals into different hybrid classes, using three runs of 500 randomly subsetting ddRAD loci. 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 [71]. 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 α, which roughly equates to the 'direction' of introgression, and β, which may be summarised as the 'rate' of introgression for a locus [25, 71]. In order to estimate these parameters from the single-SNP dataset consisting of 19,130 RAD loci, 50,000 MCMC generations with a 50% burn-in were used in *bgc*. Admixture proportions (i.e., mean *Q* values) generated by *fastSTRUCTURE* were used to assign each individual to three populations (*Brownea grandiceps*, *B. jaramilloi* and admixed). Convergence was checked for the MCMC output from *bgc* in Tracer v1.6 [72] and with the R package *coda* [73] using Geweke's diagnostic [74]. Loci with significant 'excess ancestry' were identified by ascertaining whether the 99% posterior probability estimates of the α and β parameters included zero (i.e., by identifying positive or negative non-zero estimates of the parameters). In addition, loci which were extreme 'introgression outliers' were identified for both parameters by identifying loci whose median estimates were not included in the 99% posterior probability credible intervals [71]. Further details of visualising shared variation, hybrid category assignment and *bgc* analysis are shown in Supplementary Methods 2.

Data Availability

The datasets that support the findings of this study are 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 numbers 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). The full phylogenomic sequence alignments underlie the phylogeny in Fig. 2, while the single-accession-per-species alignments and tree files underlie Fig. 3. The variant call format (VCF) and Structure files provided on Dryad underlie Fig. 4 and Fig. 5. Data are under embargo until publication, and any further data required are available from the corresponding author upon request. A reporting summary for this article is available as a Supplementary Information file.

References

1. Mayr, E. in *Systematics and the Origin of Species* (Columbia University Press, New York, New York, USA., 1942).

2. Barraclough, T. G. in *The Evolutionary Biology of Species* (Oxford University Press, Oxford, 2019).

3. Kearns, A. M. *et al*. Genomic evidence of speciation reversal in ravens. *Nat. Comm.* **9**, 906 (2018).

4. Vonlanthen, P. *et al*. Eutrophication causes speciation reversal in whitefish adaptive radiations. *Nature* **482**, 357-362 (2012).

5. Mallet, J. Hybrid speciation. *Nature* **446**, 279-283 (2007).

6. Antonelli, A. & Sanmartín, I. Why are there so many plant species in the Neotropics? *Taxon* **60**, 403-414 (2011).

7. Ehrendorfer, F. Evolutionary patterns and strategies in seed plants. *Taxon* **19**, 185-195 (1970).

8. Ashton, P. S. Speciation among tropical forest trees: some deductions in the light of recent evidence. *Biol. J. Linn. Soc.* **1**, 155-196 (1969).

9. Gentry, A. H. Neotropical floristic diversity: phytogeographical connections between Central and South America, Pleistocene climatic fluctuations, or an accident of the Andean orogeny? *Ann. Mo. Bot. Gard.* **69**, 557-593 (1982).

10. Koptur, S. Outcrossing and pollinator limitation of fruit set: breeding systems of neotropical Inga trees (Fabaceae: Mimosoideae). *Evolution* **38**, 1130-1143 (1984).

11. Scotti-Saintagne, C. *et al*. Phylogeography of a species complex of lowland Neotropical rain forest trees (Carapa, Meliaceae). *J. Biogeogr.* **40**, 676-692 (2013).

12. Cardoso, D. *et al*. Amazon plant diversity revealed by a taxonomically verified species list. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10695-10700 (2017).

13. Dexter, K. G. *et al*. Dispersal assembly of rain forest tree communities across the Amazon basin. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 2645-2650 (2017).

14. Valencia, R., Balslev, H. & Miño, G. P. Y. High tree alpha-diversity in Amazonian Ecuador. *Biodiversity & Conservation* **3**, 21-28 (1994).

15. Hopkins, R. Reinforcement in plants. *New Phytol.* **197**, 1095-1103 (2013).

16. Rieseberg, L. H. & Wendel, J. F. Introgression and its consequences in plants. *Hybrid Zones and the Evolutionary Process* **70**, 70-109 (1993).

17. Taylor, S. A. & Larson, E. L. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution* **3**, 170-177 (2019).

18. Whitney, K. D., Randell, R. A. & Rieseberg, L. H. Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytol.* **187**, 230-239 (2010).

19. Suarez-Gonzalez, A., Lexer, C. & Cronk, Q. C. B. Adaptive introgression: a plant perspective. *Biol. Lett.* **14**, 10.1098/rsbl.2017.0688 (2018).

20. Marques, D. A., Meier, J. I. & Seehausen, O. A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution* **34**, 531-544 (2019).

21. Lamichhaney, S. *et al*. Rapid hybrid speciation in Darwin's finches. *Science* **359**, 224-228 (2018).

22. Rieseberg, L. H. *et al*. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* **301**, 1211-1216 (2003).

23. Meier, J. I. *et al*. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications* **8**, 14363 (2017).

24. Payseur, B. A. Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* **10**, 806-820 (2010).

25. Gompert, Z., Parchman, T. L. & Buerkle, C. A. Genomics of isolation in hybrids. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **367**, 439-450 (2012).

26. Hamilton, J. A. & Aitken, S. N. Genetic and morphological structure of a spruce hybrid (*Picea sitchensis* × *P. glauca*) zone along a climatic gradient. *Am. J. Bot.* **100**, 1651-1662 (2013).

27. Sullivan, A. R., Owusu, S. A., Weber, J. A., Hipp, A. L. & Gailing, O. Hybridization and divergence in multi-species oak (*Quercus*) communities. *Bot. J. Linn. Soc.* **181**, 99-114 (2016).

28. Abbott, R. J. *et al*. Hybridization and speciation. *J. Evol. Biol.* **26**, 229-246 (2013).

29. Seehausen, O. *et al*. Genomics and the origin of species. *Nature Reviews Genetics* **15**, 176-192 (2014).

30. Schley, R. J. *et al*. Is Amazonia a 'museum'for Neotropical trees? The evolution of the Brownea clade (Detarioideae, Leguminosae). *Mol. Phylogenet. Evol.* **126**, 279-292 (2018).

31. Crawford, T. & Nelson, E. Irish Horticulturists. I: WH Crawford. *Garden History* **7**, 23-26 (1979).

32. Backer, C. A. in *Schoolflora voor Java* 418 (Visser & Co., Weltevreden, Java, Dutch East Indies, 1911).

33. Pérez, Á J., Klitgård, B. B., Saslis-Lagoudakis, C. & Valencia, R. *Brownea jaramilloi* (Leguminosae: Caesalpinioideae), a new, over-looked species endemic to the Ecuadorian Amazon . *Kew Bull.* **68**, 157-162 (2013).

34. Klitgaard, B. B. Ecuadorian *Brownea* and *Browneopsis* (Leguminosae-Caesalpinioideae): taxonomy, palynology, and morphology. *Nord. J. Bot.* **11**, 433-449 (1991).

35. ter Steege, H. *et al*. Hyperdominance in the Amazonian tree flora. *Science* **342**, 1243092 (2013).

36. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).

37. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).

38. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254-267 (2005).

39. Raj, A., Stephens, M. & Pritchard, J. K. fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573-589 (2014).

40. Anderson, E. C. & Thompson, E. A. A model-based method for identifying species hybrids using multilocus genetic data. *Genetics* **160**, 1217-1229 (2002).

41. Burbrink, F. T. & Gehara, M. The biogeography of deep time phylogenetic reticulation. *Syst. Biol.* **67**, 743-744 (2018).

42. Pennington, R. T. & Lavin, M. The contrasting nature of woody plant species in different neotropical forest biomes reflects differences in ecological stability. *New Phytol.* **210**, 25-37 (2016).

43. Lepais, O. & Gerber, S. Reproductive patterns shape introgression dynamics and species succession within the European white oak species complex. *Evolution* **65**, 156-170 (2011).

44. Morales-Briones, D. F., Liston, A. & Tank, D. C. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus Lachemilla (Rosaceae). *New Phytol.* **218**, 1668-1684 (2018).

45. Sweigart, A. L. & Willis, J. H. Patterns of nucleotide diversity in two species of *Mimulus* are affected by mating system and asymmetric introgression. *Evolution* **57**, 2490-2506 (2003).

46. Buggs, R. J. & Pannell, J. R. Rapid displacement of a monoecious plant lineage is due to pollen swamping by a dioecious relative. *Current Biology* **16**, 996-1000 (2006).

47. Petit, R. J., Bodénès, C., Ducousso, A., Roussel, G. & Kremer, A. Hybridization as a mechanism of invasion in oaks. *New Phytol.* **161**, 151-164 (2004).

48. Balao, F., Casimiro-Soriguer, R., García-Castaño, J. L., Terrab, A. & Talavera, S. Big thistle eats the little thistle: does unidirectional introgressive hybridization endanger the conservation of *Onopordum hinojense*? *New Phytol.* **206**, 448-458 (2015).

49. Yang, W. *et al*. Genomic evidence for asymmetric introgression by sexual selection in the common wall lizard. *Mol. Ecol.* **27**, 4213-4224 (2018).

50. Scascitelli, M. *et al*. Genome scan of hybridizing sunflowers from Texas (*Helianthus annuus* and *H. debilis*) reveals asymmetric patterns of introgression and small islands of genomic differentiation. *Mol. Ecol.* **19**, 521-541 (2010).

51. Arnold, M. L., Tang, S., Knapp, S. J. & Martin, N. H. Asymmetric introgressive hybridization among Louisiana Iris species. *Genes* **1**, 9-22 (2010).

52. Mallet, J. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* **20**, 229-237 (2005).

53. Kenzo, T. *et al*. Overlapping flowering periods among *Shorea* species and high growth performance of hybrid seedlings promote hybridization and introgression in a tropical rainforest of Singapore. *For. Ecol. Manage.* **435**, 38-44 (2019).

54. Kamiya, K. *et al*. Morphological and molecular evidence of natural hybridization in *Shorea* (Dipterocarpaceae). *Tree Genetics & Genomes* **7**, 297-306 (2011).

55. Lo, E. Testing hybridization hypotheses and evaluating the evolutionary potential of hybrids in mangrove plant species. *J. Evol. Biol.* **23**, 2249-2261 (2010).

56. Solís-Lemus, C., Bastide, P. & Ané, C. PhyloNetworks: a package for phylogenetic networks. *Mol. Biol. Evol.* **34**, 3292-3298 (2017).

57. Legume Phylogeny Working Group. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). *Taxon* **66**, 44-77 (2017).

58. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc. (2010).

59. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

60. Johnson, M. G. *et al*. HybPiper: Extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *App. Plant. Sci.* **4**, 1600016 (2016).

61. Python Software Foundation. Python Language Reference, version 2.7. *http://www.python.org* (2010).

62. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772-780 (2013).

63. Smith, S. A., Moore, M. J., Brown, J. W. & Yang, Y. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* **15**, Published online: DOI 10.1186/s12862-015-0423-0 (2015).

64. Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM Rev* **59**, 65-98 (2017).

65. Stenz, N. W., Larget, B., Baum, D. A. & Ané, C. Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. *Syst. Biol.* **64**, 809-823 (2015).

66. R Development Core Team. R: A language and environment for statistical computing. Available at: http://www.R-project.org/. (2013).

67. Ho, L. S. T. & Ané, C. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst. Biol.* **63**, 397-408 (2014).

68. 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).

69. 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).

70. Wickham, H. in *ggplot2: elegant graphics for data analysis* (Springer, 2016).

71. Gompert, Z. & Buerkle, C. A. Bayesian estimation of genomic clines. *Mol. Ecol.* **20**, 2111-2127 (2011).

72. Rambaut, A., Suchard, M. A., Xie, D. & Drummond, A. J. Tracer. *http://tree.bio.ed.ac.uk/software/tracer/* **1.6** (2015).

73. Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R news* **6**, 7-11 (2006).

74. Geweke, J. in *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* (Federal Reserve Bank of Minneapolis, Research Department, Minneapolis, MN, USA, 1991).

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. Thank you to Dario Ojeda for baits and associated help, to Laszlo Csiba, Penny Malakasi, Niroshini Epitawalage, Dion Devey, Robyn Cowan and Steven Dodsworth for help in the laboratory, and to Gregor Gilfillan at the Norwegian sequencing centre for help with ddRAD sequencing. Thank you to Andres Melo Burbano for his help during fieldwork, as well as to Renato Valencia and Álvaro Pérez for helping to coordinate fieldwork in Ecuador. Thanks to Colin Hughes, Jeff Doyle and Matteo Fumagalli for comments on analytical methods, and to 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 authorised 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

R.J.S. conceived this study and performed all analyses, supervised by T.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.l.E. R.J.S. wrote the first draft of the manuscript with contributions from R.T.P., O.A.P.E., A.J.H., M.d.l.E., I.L., T.B., F.F. and B.K.

Competing Interests

The authors declare no competing financial interest.