



Resolving the Speciation Patterns and Evolutionary History of the Intercontinental Disjunct Genus *Corylus* (Betulaceae) Using Genome-Wide SNPs

Zhen Yang, Tian-Tian Zhao*, Qing-Hua Ma, Li-Song Liang and Gui-Xi Wang*

Key Laboratory of Tree Breeding and Cultivation of the State Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China

OPEN ACCESS

Edited by:

Octavio Salgueiro Paulo,
Universidade de Lisboa, Portugal

Reviewed by:

Nakatada Wachi,
University of the Ryukyus, Japan
Isabel Marques,
University of British Columbia,
Canada

*Correspondence:

Tian-Tian Zhao
Zhaotian1984@163.com
Gui-Xi Wang
wanggx0114@126.com

Specialty section:

This article was submitted to
Evolutionary and Population Genetics,
a section of the journal
Frontiers in Plant Science

Received: 23 December 2017

Accepted: 31 August 2018

Published: 25 October 2018

Citation:

Yang Z, Zhao T-T, Ma Q-H, Liang L-S
and Wang G-X (2018) Resolving
the Speciation Patterns
and Evolutionary History of the
Intercontinental Disjunct Genus
Corylus (Betulaceae) Using
Genome-Wide SNPs.
Front. Plant Sci. 9:1386.
doi: 10.3389/fpls.2018.01386

Understanding the underlying mechanisms of species origin, divergence, and distribution patterns of the intercontinental disjunct taxa has long fascinated botanists. Based on 4,894 genome-wide single-nucleotide polymorphism dataset, we present a molecular phylogenetic reconstruction of genus *Corylus* (Betulaceae), which have a disjunct distribution between Eurasia and North America (NA). The aim is to explore the speciation patterns and evolutionary relationships of *Corylus* species by establishing a general phylogenetic framework with extensive sampling. Both the molecular phylogeny inferred from recombination-free dataset and structure analysis support the division of *Corylus* into four major clades (A–D). Recombination tests and hybridization detection reveal extensive recombination and hybridization events among different clades, which have potentially influenced the speciation process of *Corylus*. Divergence time estimation indicates that recent common ancestor (MRCA) of *Corylus* occurred in late Eocene (~36.38 Ma) and subsequent rapid diversification began during Miocene. Ancestral area reconstruction shows that *Corylus* originated from southwest China. The arrival of two clades (Clades B and C) to NA was well supported by the long distance dispersal crossing the Bering land bridge. The Himalayas, European-Mediterranean area, and other distribution regions are primarily the recipients of dispersal taxa. Vicariance after dispersal plays an important role in speciation.

Keywords: *Corylus*, speciation, recombination, hybridization, divergence time estimation, ancestral area reconstruction, genome-wide SNPs

INTRODUCTION

In the Northern Hemisphere, intercontinental disjunction of related plant species among eastern Asia (EA), Europe, and North America (NA) has long fascinated botanists and biogeographers (Tiffney and Manchester, 2001; Xiang and Soltis, 2001; Donoghue and Smith, 2004; Wen and Ickert-bond, 2009; Xiang et al., 2015). The disjunction patterns have been utilized to grasp the histories of plant dispersal between continents as well as allopatric speciation (Boufford and Spongberg, 1983; Wen, 1999; Kelchner and Bamboo Phylogeny Group, 2013). Multiple origins and complex evolutionary patterns of this intercontinental disjunction have been discussed based on fossils, molecular, and geologic evidence

(Wen, 1999; Xiang et al., 2004). Despite various interpretations for the disjunction based on different taxa, it has been generally recognized that climatic fluctuations over the Cenozoic and two intercontinental land bridges, i.e., the North Atlantic Land Bridge (NALB) and the Bering land bridge (BLB), have played important roles in shaping current disjunctions of the Northern Hemisphere flora. Nevertheless, due to the complex biotic responses to diverse abiotic factors in the Northern Hemisphere, there still remains much to explore about the evolutionary history of the disjunction patterns and the underlying mechanisms of species diversification.

Understanding the speciation of diverse lineages or species-rich communities is of great interest in biological sciences. The speciation mechanisms may be associated with various factors such as probable gene mutations, potential recombination, and hybridization phenomena, and a series of dispersal and vicariance events. Mutation and gene recombination provide the original impetus for biological evolution, which is the intrinsic factor of speciation, while selection will externally retain dominant species that can be well suitable for certain ecological conditions. Especially, closely related species may also hybridize if reproductive isolation is incomplete, leading to diverse possible outcomes such as the decline or extinction of one or both parental species through genetic or demographic swamping (Holt and Gomulkiewicz, 1997; Wolf et al., 2001), establishment of recombinant species (Livingstone and Rieseberg, 2004), or the transfer of adaptive alleles (Arnold et al., 2016). Furthermore, more studies indicate that speciation has resulted from dispersal or a complex mix involving both vicariance and dispersal (Mao et al., 2010; Migliore et al., 2012; Chen et al., 2014). However, comprehensive analyses including all these aspects are very few.

Corylus L. (Betulaceae), the hazelnut genus, provides an ideal model for studying the evolution of intercontinental disjunctions in the Northern Hemisphere, as well as the diversification within EA. The genus consists of approximately 15–20 species disjunctly distributed in major areas of the Northern Hemisphere, with high species diversity in EA (especially in China). About 10 species occur in China, 1 in Korea and Japan, 3 in NA, 1 in the Himalayas, and 2 in Europe and the Mediterranean regions. Although the distribution of *Corylus* species shows a noticeable disjunct pattern, the biogeographical study by Whitcher and Wen (2001) is the only study that examined this pattern by calculating the substitution rate of the internal transcribed spacer (ITS) region. Additionally, age estimation of *Corylus* has not been conducted in spite of abundant fossil records (Crane, 1989; Chen et al., 1999). Therefore, origin of genus *Corylus* species and their biogeographic patterns have not been addressed.

The genus *Corylus* is characterized by several morphological synapomorphies, including large animal-dispersed nuts, hypogeal seed germination, and filaments that are completely divided longitudinally (Chen et al., 1999). The chromosome number of this genus is $2n = 2x = 22$ (Thompson et al., 1996). Classification in the genus has traditionally been based on morphology, especially in the husk or involucre (Krüssmann, 1976; Everett, 1982; Huxley and Griffiths, 1999). The number of *Corylus* species has varied depending on various authors. Infrageneric taxonomy has recognized two main sections or

subgenera (*Acanthochlamys* and *Corylus*); with section *Corylus* often being divided into three subsections (Furlow, 1997). While several classification treatments have been limited to taxa in a regional scale (Li and Cheng, 1979; Liang and Zhang, 1988; Furlow, 1997), even among classifications treating the same species, inclusion of taxa within each section or subgenus has varied significantly. Species identification is also controversial in *Corylus*. Particularly, two species complexes have been subjected to different taxonomic interpretations: *Corylus heterophylla* Fisch. complex and the *C. cornuta* Marsh. complex (Li and Cheng, 1979). The *C. heterophylla* complex distributes widely in EA, including three leafy-husked shrubs: *C. heterophylla* Fisch., *C. kweichowensis* Hu, and *C. yunnanensis* Fisch. The *C. cornuta* complex is a group of EA and NA taxa which contain four bristle-husked shrubs: *C. cornuta* Marshall, *C. californica* Marshall, *C. sieboldiana* Blume, and *C. mandshurica* Maxim. The species within each complex have been variously lumped and split. Over the past decades, several molecular studies have provided important insights into the phylogeny and taxonomy of *Corylus* (Erdogan and Mehlenbacher, 2000; Whitcher and Wen, 2001; Bassil et al., 2013). However, these above studies are still failed to reach an ideal result partly because of the incomplete taxa sampling and partly due to the low resolution in species delimitation.

In this study, we sampled extensively from across the known geographical range of *Corylus* and performed multiple analyses based on the genome-wide single-nucleotide polymorphism (SNP) dataset. Our aims are (1) to establish a robust molecular phylogeny and reveal the evolutionary relationships of *Corylus*; (2) to test the potential recombination and hybridization events that might affect the speciation of *Corylus*; and (3) to estimate the divergence time and history biogeography of *Corylus*.

MATERIALS AND METHODS

Study System

Members of genus *Corylus* are perennial shrubs or trees that vary most notably in the tree trunk, leaf, and involucre characteristics (Figure 1). Approximately 16–20 species occur across Eurasia and NA, with East Asia especially China as the main species enrichment zone. We employed the accepted names from *The Plant List*¹ and *Flora of China* (Editorial Board of the Flora of China of Chinese Academy of Sciences, 2013), and replenished it by referring to other relevant floristic treatments: *Flora Europaea* (Tutin and Walters, 1993), *Flora of NA* (Furlow, 1997), *Flora of the USSR* (Bobrov, 1936), and *Flora of Japan* (Ohwi, 1965). Simultaneously, we consulted the herbarium specimens of Institute of Botany, Chinese Academy of Sciences. For further correction, we compared the existing identification with that of *World Checklist of Selected Plant Families* (WCSP) so as to filter synonymy or unresolved species names. Here, several scenarios for the treatments of ambiguous species were described. (1) Despite of an accepted name in *The Plant List*, *C. colchica* has

¹<http://www.theplantlist.org>



FIGURE 1 | Phenotypic characters of several representative *Corylus* species. **(A–G)** Tree forms. **(H–O)** Leaf shapes. **(P–V)** Husk shapes. **(W,X)** Husk hairs. **(Y)** Stem.

few records (Bobrov, 1936; Govaerts and Frodin, 1998) and remains a bit of uncertainty (Holstein et al., 2018). Therefore, this alleged species is not considered to be included in our samples. (2) *C. maxima*, one of the most mysterious species, has been widely reported as a wild species occurring in European-Mediterranean area (Bobrov, 1936; Tutin and Walters, 1993), however, relevant phylogenetic analysis (Whitcher and Wen, 2001) has revealed that it is probably a variety of *C. avellana*. Thus, *C. maxima* was covered and represented by *C. avellana* in this study. (3) Although the two Chinese species *C. potaninii* (Govaerts and Frodin, 1998) and *C. wulingensis* (Govaerts and Frodin, 1998) are both designated as accepted names in *The Plant List*, they are no longer recognized as distinct *Corylus*

species and not mentioned in *Flora of China* (Editorial Board of the Flora of China of Chinese Academy of Sciences, 2013). Recent classification in China has confirmed them as the same species with the name *C. kweichowensis* (Liang and Zhang, 1988; Editorial Board of the Flora of China of Chinese Academy of Sciences, 2013). (4) As for the synonym, *C. mandshurica* and *C. californica* are separately regarded as the variety of *C. sieboldiana* var. *C. mandshurica* (Ohwi, 1965) and *C. cornuta* subsp. *californica* (Furrow, 1997) by some taxonomists, but as distinct species by others (Bassil et al., 2013). In view of the significant geographical isolation and covering more taxa, we treat them as four distinct species so as to conduct a comprehensive species analysis.

Taxon Sampling and DNA Extraction

Based on the above pre-treatments, we finally chose as inclusive as possible the 17 taxa which represented the most complete examination of the genus *Corylus* to date. In fact, our team has initiated a comprehensive investigation and collection for *Corylus* germplasm resources since 2010 by cooperating with relevant research institutes in China, America, Turkey, and Netherlands. Therefore, we are convenient to collect all the 17 extant *Corylus* taxa, and the species delimitation work received great support from these specialists. According to our survey, nearly all the *Corylus* resources are under wild or semi-wild state except for *C. avellana* which has numerous of cultivars in the market. Consequently, various nature nurseries were *in situ* established by local forestry institutions to protect these wild resources without disrupting their genetic diversity. Ten (eight species and two varieties) of the 17 taxa were taken from the natural populations that covered their distribution ranges in China. The remaining seven taxa were friendly provided by the herbarium specimens deposited in countries and regions of Europe, NA, and the Himalayas. It is worth noting that these specimens were also collected from the representative natural populations across their distribution areas (Table 1). Overall, a total of 45 specimens were collected, of which, the ingroup included 42 samples representing 17 *Corylus* species (varieties), while the other three samples of *Ostryopsis davidiana* Decne were chosen as the outgroup. Voucher specimens were deposited in the Economic forest research office of Research Institute of Forestry Chinese Academy of Forestry, Beijing, China.

High-Throughput Sequencing, Data Filtering, and SNP Genotyping

To generate genome-wide data to infer the evolutionary history of genus *Corylus*, we used Illumina sequencing of 2b-restriction site-associated DNA (2b-RAD). To guarantee the yield of DNA from herbarium material, genomic DNA was isolated using the modified cetyltrimethyl ammonium bromide (CTAB) protocol (Sobel and Streisfeld, 2015). The DNA concentration and purity were evaluated both with a NanoDrop-2000 spectrophotometer. The final concentration of DNA samples was diluted to 200 ng μL^{-1} . 2b-RAD libraries were then prepared for each individual using the type IIB enzyme BsaXI, followed by single-end sequencing using an Illumina HiSeq X Ten platform, according to the protocol developed by Wang et al. (2012). We used the software Stacks v. 1.35 (Catchen et al., 2011) to remove reads with low quality or uncalled bases. Errors in the restriction site sequences and barcode were checked before downstream analysis. Then, reads were aligned to the finished reference genome from *Betula nana* (EMBL accession number ERP001867; Wang et al., 2013) using SOAP v. 2.21 (Li et al., 2009). SNP genotyping was performed using the program RADtyping v1.537 (Fu et al., 2013) with the maximum-likelihood (ML) algorithm (all remaining parameters as default). The resulted SNPs were further filtered for our subsequent analysis using the population module in Stacks v. 1.35, requiring that SNPs occur in at least 80% of the individuals and had a minimum minor allele frequency of 0.01 to exclude any SNP locus found in a single heterozygote.

TABLE 1 | Details of taxon code, sample code, and sampling location of 45 individuals used in the study.

No.	Taxon code	Sample code	Location
1	Taxa 1	<i>C. colurna-1</i>	Poland
2		<i>C. colurna-2</i>	Czech Republic
3		<i>C. colurna-3</i>	Georgia
4	Taxa 2	<i>C. jacquemontii-1</i>	Pakistan
5		<i>C. jacquemontii-2</i>	Pakistan
6		<i>C. jacquemontii-3</i>	Nepal
7	Taxa 3	<i>C. fargesii-1</i>	Gansu, China
8		<i>C. fargesii-2</i>	Gansu, China
9	Taxa 4	<i>C. chinensis-1</i>	Yunnan, China
10		<i>C. chinensis-2</i>	Shanxi, China
11		<i>C. chinensis-3</i>	Shanxi, China
12	Taxa 5	<i>C. avellana-1</i>	Oregon, United States
13		<i>C. avellana-2</i>	Oregon, United States
14		<i>C. avellana-3</i>	Oregon, United States
15	Taxa 6	<i>C. americana-1</i>	Michigan, United States
16		<i>C. americana-2</i>	Oregon, United States
17		<i>C. americana-3</i>	Oregon, United States
18	Taxa 7	<i>C. yunnanensis-1</i>	Yunnan, China
19		<i>C. yunnanensis-2</i>	Yunnan, China
20		<i>C. yunnanensis-3</i>	Yunnan, China
21	Taxa 8	<i>C. kweichowensis-1</i>	Shanxi, China
22		<i>C. kweichowensis-2</i>	Shanxi, China
23	Taxa 9	<i>C. kweichowensis</i> var. <i>brevipes</i>	Jiangxi, China
24	Taxa 10	<i>C. heterophylla-1</i>	Hebei, China
25		<i>C. heterophylla-2</i>	Jilin, China
26	Taxa 11	<i>C. sieboldiana-1</i>	Japan
27		<i>C. sieboldiana-2</i>	Japan
28		<i>C. sieboldiana-3</i>	Korea
29	Taxa 12	<i>C. mandshurica-1</i>	Beijing, China
30		<i>C. mandshurica-2</i>	Hebei, China
31	Taxa 13	<i>C. californica-1</i>	Oregon, United States
32		<i>C. californica-2</i>	Oregon, United States
33		<i>C. californica-3</i>	Oregon, United States
34	Taxa 14	<i>C. cornuta-1</i>	Minnesota, United States
35		<i>C. cornuta-2</i>	Minnesota, United States
36		<i>C. cornuta-3</i>	New York, United States
37	Taxa 15	<i>C. ferox-1</i>	Yunnan, China
38		<i>C. ferox-2</i>	Shanxi, China
39	Taxa 16	<i>C. ferox</i> var. <i>thibetica-1</i>	Shanxi, China
40		<i>C. ferox</i> var. <i>thibetica-2</i>	Shanxi, China
41	Taxa 17	<i>C. wangii-1</i>	Yunnan, China
42		<i>C. wangii-2</i>	Yunnan, China
43	Taxa 18	<i>Ostryopsis davidiana-1</i>	Neimenggu, China
44		<i>Ostryopsis davidiana-2</i>	Neimenggu, China
45		<i>Ostryopsis davidiana-3</i>	Neimenggu, China

Phylogenetic Analysis

To infer relationships among samples, we performed phylogenetic analyses with concatenated SNP dataset using both the ML and Bayesian Inference (BI) methods. The optimal substitution models for the ML and BI phylogenetic analyses

were determined by ModelFinder program (Kalyaanamoorthy et al., 2017) using the Bayesian information criterion (BIC), as implemented in IQ-TREE (Nguyen et al., 2014). The ML analysis was conducted with IQ-TREE using 1000 replicates of ultrafast bootstrapping (UFBoot: Minh et al., 2013) and 1000 bootstrap replicates of the Shimodaira/Hasegawa approximate likelihood-ratio test (SH-aLRT: Guindon et al., 2010). The BI analysis was performed in MrBayes 3.2 (Ronquist et al., 2012) by running for 100,000 generations and sampling every 100 generations with the selected evolutionary model. The run was not finished until the average standard deviation of split frequencies was lower than 0.01 in all cases. The first 25% of the trees were discarded as burn-in, and the remaining trees were used to construct a 50% majority-rule consensus tree and estimate the Bayesian posterior probabilities. Trees were visualized and edited in FigTree 1.4.0 (Rambaut, 2012).

Recombination Tests and Phylogeny Reassessment

Recombination between nucleotide sequences is a major process influencing the evolution of most species. To track the potential recombination events between two introgressed ecotypes and their influence on phylogenetic inference, unguided tests for recombination were performed in the RDP4 program (Martin et al., 2015). RDP4 executes a hidden Markov model to estimate the breakpoint positions once a recombination event is identified. In our analysis, recombination tests were conducted with seven implemented algorithms using the RDP (Martin and Rybicki, 2000), GENECONV (Padidam et al., 1999), MaxChi (Smith, 1992), Chimaera (Posada and Crandall, 2001), BootScan (Martin et al., 2005), 3Seq (Boni et al., 2007), and SiScan (Gibbs et al., 2000) methods. We tested the SNP dataset alignment for recombination, with Bonferroni corrections applied to set the family-wise error rate to 0.05. Only results that were supported by at least four of the seven approaches were accepted.

Inference of phylogeny using genome-wide SNPs can be severely distorted by recombination events. Based on the recombination tests, we did detect three obvious recombination signals which may probably result in unreliable topologies. Therefore, we further conducted a modified phylogeny (ML and BI) using the recombination-free dataset to check the topological changes. The recombination-free dataset was generated from RDP4 and the method for phylogenetic analysis was the same as above.

Admixture Analysis and Hybridization Detection

To test for admixture and to infer potential hybridization events between different clades, two Bayesian clustering procedures were applied. Firstly, the genetic structure was analyzed using software STRUCTURE 2.3.4 (Pritchard et al., 2000) to infer patterns of ancestry within *Corylus* species. Compared with phylogenetic methods, structure analysis reveals shared variation among inferred subgroups, which could result from admixture and hybridization. An admixture model with correlated allele frequencies was applied to identify the putative number of

subgroups (K). Because the phylogenetic analysis revealed four major clades (Clades A–D) in the ingroup, we thus set the targeted K from 2 to 10, with 10 independent simulations for each K (100,000 burn-ins and 100,000 iterations). Then, Structure Harvester was used to capture the true number of populations as described by Earl (2012). Assignment to clusters was then compared to known clade composition in phylogenetic analysis.

Furthermore, the Bayesian model-based method implemented in NewHybrids v 1.1 (Anderson and Thompson, 2002) was further applied to compute the posterior probability that an individual belongs to distinct genotype frequency classes (two parents, F1 and F2 hybrids, and first generation hybrid backcrosses) corresponding to hybrid categories. We used 245 SNPs from our 2b-RAD dataset which were filtered the set of loci using the GenAlex 6.5 (Peakall and Smouse, 2012) to include SNPs with a minor allele frequency > 0.2. NewHybrids was executed with four replicate runs of 100,000 sweeps and a burn-in of 100,000 sweeps with default genotype categories. The Jeffrey's priors were chosen to downweight the influence of an allele that might be rare in one species and absent in the other.

Divergence Time Estimation

Molecular dating analysis was performed in BEAST 2.5 (Bouckaert et al., 2014), using an uncorrelated relaxed clock and a Yule speciation process to estimate the divergence time of *Corylus* at the interspecific level. This analysis involved 43 individuals representing distinctly separated genetic lineages or different species, of which *O. davidiana* was chosen as outgroup. For divergence time estimation, the general time-reversible nucleotide substitution model with among-site rate variation modeled with a gamma distribution (GTR + Γ) was selected. Based on previous biogeographical studies of order Fagales (Zhang, 2014), and two fossil records of *Corylus* species (Takhtajan, 1982; Wolfe and Wehr, 1987), three calibrations were selected in the analysis: (1) The root of the tree was calibrated with the stem age of *Ostryopsis*, using a normal distribution with a mean date of 37.2 Ma and a standard deviation of 4.0 (Zhang, 2014); (2) the stem age of *Corylus* was set to 41.7 Ma based on the fossil record (Wolfe and Wehr, 1987), using a normal distribution with standard deviation of 3.8 Ma; (3) the divergence date of the subsection *Colurnae* (Clade D) is based on the clear fossil record, which is a fossil fruit similar to modern *C. colurna* or *C. chinensis*, supporting the existence of this clade with a mean date 9.82 Ma (8.74–10.9 Ma) (Takhtajan, 1982). The input file for BEAST2, with all the parameters and priors, was set up in BEAUti 2.5 using the Bayesian method based on the Markov Chain Monte Carlo (MCMC) algorithm (Bouckaert et al., 2014). Molecular dating analysis was run for a total of 100 million generations with a sampling frequency of 1,000 generations. The adequacy of parameters was checked using Tracer v.1.6, noting effective sample size (ESS) values > 200. A 25% burn-in was applied in TreeAnnotator 2.1.2 (Rambaut and Drummond, 2014), and the posterior sample estimates of the trees were summarized and combined to produce a consensus maximum clade credibility tree. Finally, FigTree 1.4 (Rambaut, 2012) was used to visualize the best molecular phylogeny and the 95% highest posterior density (HPD) for each node.

Ancestral Area Reconstruction

To reconstruct the broad-scale biogeographical history of *Corylus*, we coded the distribution of each extant species as a character with eight states according to floristic division proposed by Zhang et al. (2005) and Bassil et al. (2013): A, Northeast Asia; B, Qinling Mountains and Central Plains of China; C, Central and East China; D, southwestern China; E, the Himalayas; F, European-Mediterranean Area; G, eastern NA; H, western NA. Ancestral area reconstruction and the estimation of the spatial patterns of geographic diversification within *Corylus* were inferred using the Bayesian Binary MCMC (BBM) method, as implemented in RASP 4.0 (Yu et al., 2015). The input file for RASP came from the post burn-in trees from the interspecific BEAST analysis. Besides, a condensed tree used for mapping the ancestral distribution on each node was generated from TreeAnnotator. The BBM analysis was run under the fixed state frequencies model (Jukes–Cantor) with equal among-site rate variation for 2 million generations, 10 chains each, and 2 parallel runs. The number of maximum areas was maintained at four.

RESULTS

SNP Genotyping

The sequencing of 45 2b-RAD libraries generated 255,951,619 raw reads (average 5,687,813 raw reads per sample). The average coverage of all genomes was 57.24 \times . After trimming the barcode, cleaning, and filtering out the low-quality reads, we obtained a total of 204,761,295 clean reads (average 4,550,251 clean reads), of which 181,770,561 were found to contain enzyme recognition sites (average 4,039,346 enzymes per sample). On average, the ratio of enzymes to clean reads in the sequencing libraries was over 74.8%, suggesting the high quality of the 45 libraries. Overall, an average of 60.40% of the high-quality reads for each sample was uniquely mapped onto the reference genome (Supplementary Table S1). Finally, a total of 4,894 SNPs were genotyped and used for subsequent analyses. The RAD data have been submitted to the Sequence Read Archive (SRA) database in the NCBI, under accession numbers SAMN09464508–SAMN09464552.

Phylogenetic Analysis

Overall, phylogenetic trees inferred from ML and BI methods showed a highly consistent topology even at small branch nodes. Here, we only displayed the ML phylogeny inferred from TVMe + R2 model, with ultrafast bootstrapping (UFBoot > 70%) and Bayesian posterior probability (PP > 95%) values displayed above branches (Figure 2). Both phylogenetic trees identified four well-supported clades (A–D) in *Corylus* and resolved the phylogenetic relationships among the major clades. Clade A (UFboot/PP: 100/1) included three ancient *Corylus* species endemic to China: *C. ferox*, *C. ferox* var. *thibetica*, and *C. wangii*. *C. ferox* and its variety *C. ferox* var. *thibetica* clustered into a common subclade, with *C. wangii* formed its sister group. Clade B (100/1), a paraphyletic clade, was composed of four shrub species that disjunctively distributed between Northeast

Asia and NA. In this clade, *C. cornuta* and *C. californica*, and *C. sieboldiana* and *C. mandshurica* formed two separate subclades, respectively. Clade C (99/1) was a species complex consisted of five morphologically similar shrubs: *C. americana*, *C. heterophylla*, *C. yunnanensis*, *C. kweichowensis*, and its variety *C. kweichowensis* var. *brevipes*. Of these species, four Chinese species (*C. heterophylla*, *C. yunnanensis*, *C. kweichowensis*, and its variety) showed a closer affinity than that with *C. americana* of American origin. Clade D (73/1) was a multi-origin group of five geographically isolated species, including *C. colurna* and *C. avellana* from European-Mediterranean region, *C. jacquemontii* from the Himalayas, and *C. fargesii* and *C. chinensis* from China. Notably, this clade was moderately supported by ultrafast bootstrap approximation.

Recombination Tests and Phylogeny Reassessment

Three putative recombination events were identified by at least four of the seven methods in RDP4 (Table 2). Both the first two recombination events were discovered to occur in Clade B (Figures 1, 2), with all the individuals involved as recombinants. The potential major and minor parents in the first recombination event were appointed to *C. kweichowensis* var. *brevipes* in Clade C and *C. ferox* var. *thibetica* in Clade A. ML breakpoints of these recombinants were roughly located within the 2,100–2,500 nucleotide regions (Figure 3). The second recombination event was highly supported by six of the seven methods (Table 2). An approximately 700 bp recombinant segments originated from the major parent *ferox* var. *thibetica* (Clade A) and the minor parent *C. jacquemontii* (Clade D), a species distributed in the Himalayas. The breakpoint of each recombinant was displayed to be seated within the 3,800–4,300 nucleotide regions. A third recombination event was found to occur in Clade D (Figures 1, 2), with three of the five species (*C. chinensis*, *C. avellana*, and *C. jacquemontii*) being related to recombinants (Table 2 and Figure 3). A small recombinant segments (~150 bp) may have originated from two unknown parents similar to *C. ferox* var. *thibetica* (Clade A) and *C. heterophylla* (Clade C), respectively.

In order to assess the influences of recombination signals on tree topology, an additional phylogenetic analysis was conducted using the recombination-free data (Figure 4A). Compared to prior phylogenetic tree (Figure 2), the newly generated tree showed a marked change by transferring *C. avellana* from Clade D to Clade C. This change was closely related to these recombination events, especially the third one, from which *C. avellana* may have obtained recombinant fragments shared by species in Clade D. However, the species composition and topology of other clades remained unchanged.

Admixture Analysis and Hybridization Detection

All individuals were further assessed for genetic stratification using the STRUCTURE program. SNP data were analyzed with the possible clustering number (*K*) ranging from 2 to 10. The ΔK showed a clear maximum for *K* = 6 (Supplementary Figure S1 and Figure 1), indicating that all individuals (including

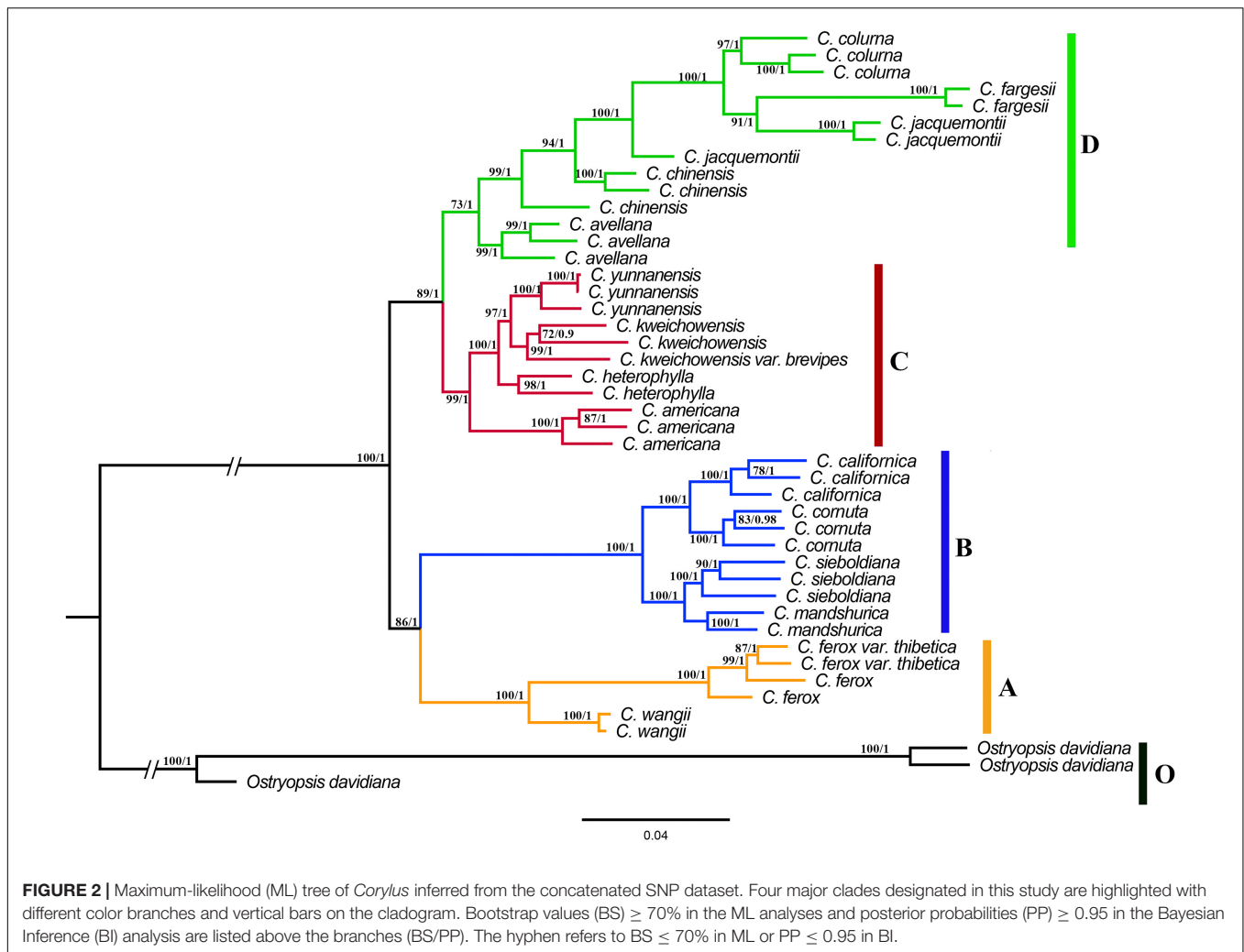


TABLE 2 | Bonferroni corrected *P*-values for the three recombination events detected in Clades B and D.

Recombination tests (methods)	Clade B-event 1 (<i>p</i> -value)	Clade B-event 2 (<i>p</i> -value)	Clade D-event 3 (<i>p</i> -value)
RDP	2.10×10^{-01}	2.89×10^{-01}	5.06×10^{-03}
GENECOV	–	–	–
BootScan	1.11×10^{-04}	2.72×10^{-03}	1.94×10^{-02}
MaxChi	5.38×10^{-03}	1.73×10^{-02}	–
Chimaera	3.41×10^{-02}	8.16×10^{-03}	4.76×10^{-02}
SiScan	1.08×10^{-02}	6.75×10^{-04}	–
3Seq	–	1.39×10^{-02}	2.03×10^{-02}

outgroup) could be classified into six optimal subgroups. From the structure plot (Figure 4B), we identified five clear subgroups which corresponded consistently with five clades (A–D and outgroup) of the recombination-free phylogeny (Figure 4A). Simultaneously, we also captured an inconspicuous but extensively existed subgroup, suggesting a complex pattern of introgression and admixture among different clades. Particularly,

C. chinensis and *C. wangii* seem to have originated from the hybridization between Clades C and D, and Clades C and A, respectively (Figures 4A,B).

Using the subset of 245 SNPs with a minor allele frequency > 0.2 , genotype frequency classes of each individual were calculated by NewHybrids (Supplementary Table S2 and Table 2). With a probability of 1, all the 11 individuals in Clade B designated by STRUCTURE were assigned as the pure parent 1, while the 25 individuals in Clade C + Clade D were classified as another pure parent 2. Interestingly, NewHybrids identified no individuals to be F1 hybrids, while four individuals of *C. ferox* and *C. ferox* var. *thibetica* in Clade A were tagged as F2 with a probability of 1. Furthermore, two individuals of *C. wangii* were classified by NewHybrids as backcross hybrids (F2 \times pure parent 2). It is noteworthy that the majority of these hybrids are from the same Clade A, indicating that the species in this clade are probably more easily to hybrid with species of other clades.

Divergence Time Estimation

The tree topology recovered from the molecular dating analysis (Figure 5) was identical to those inferred from the ML analysis

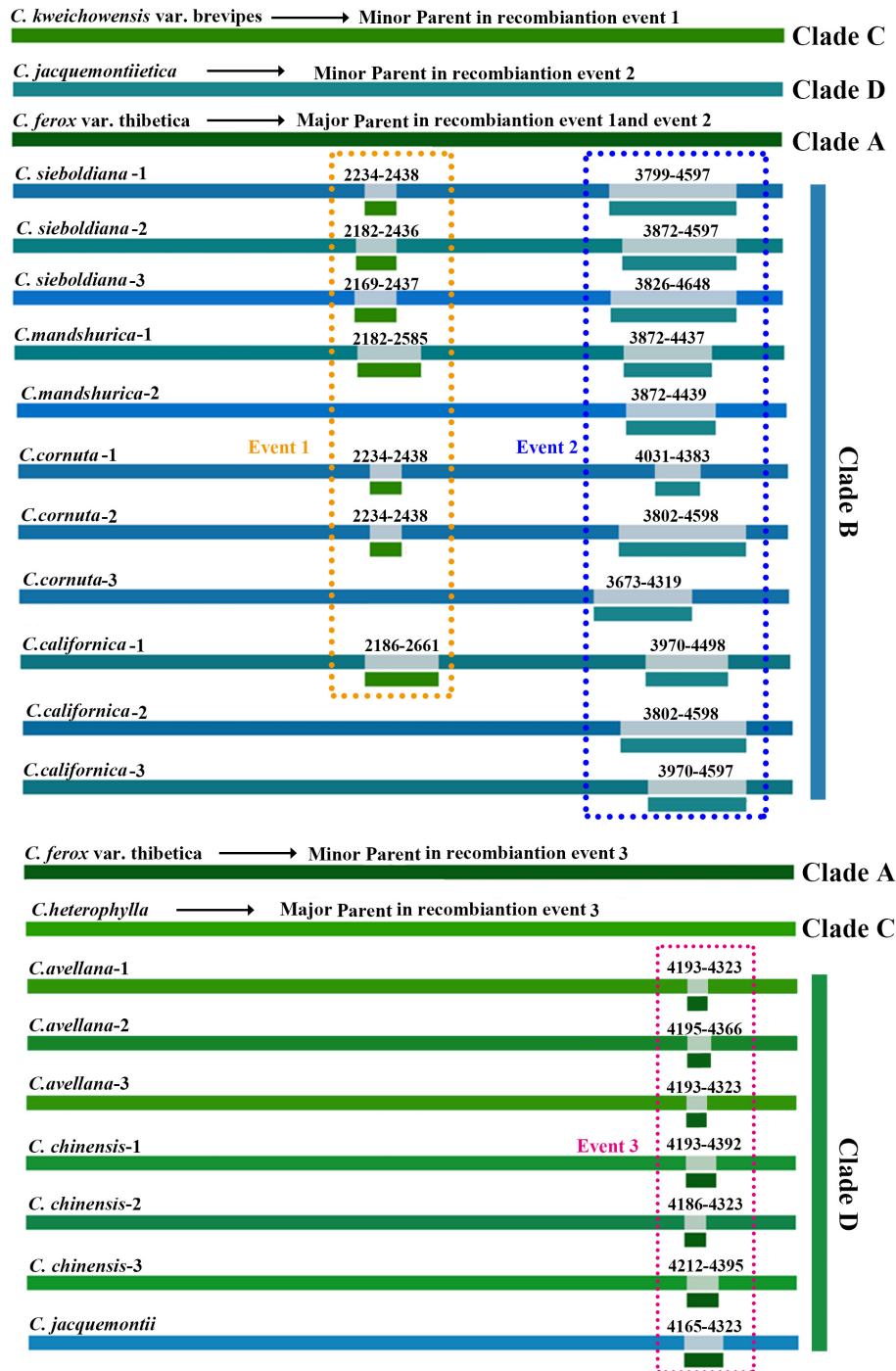
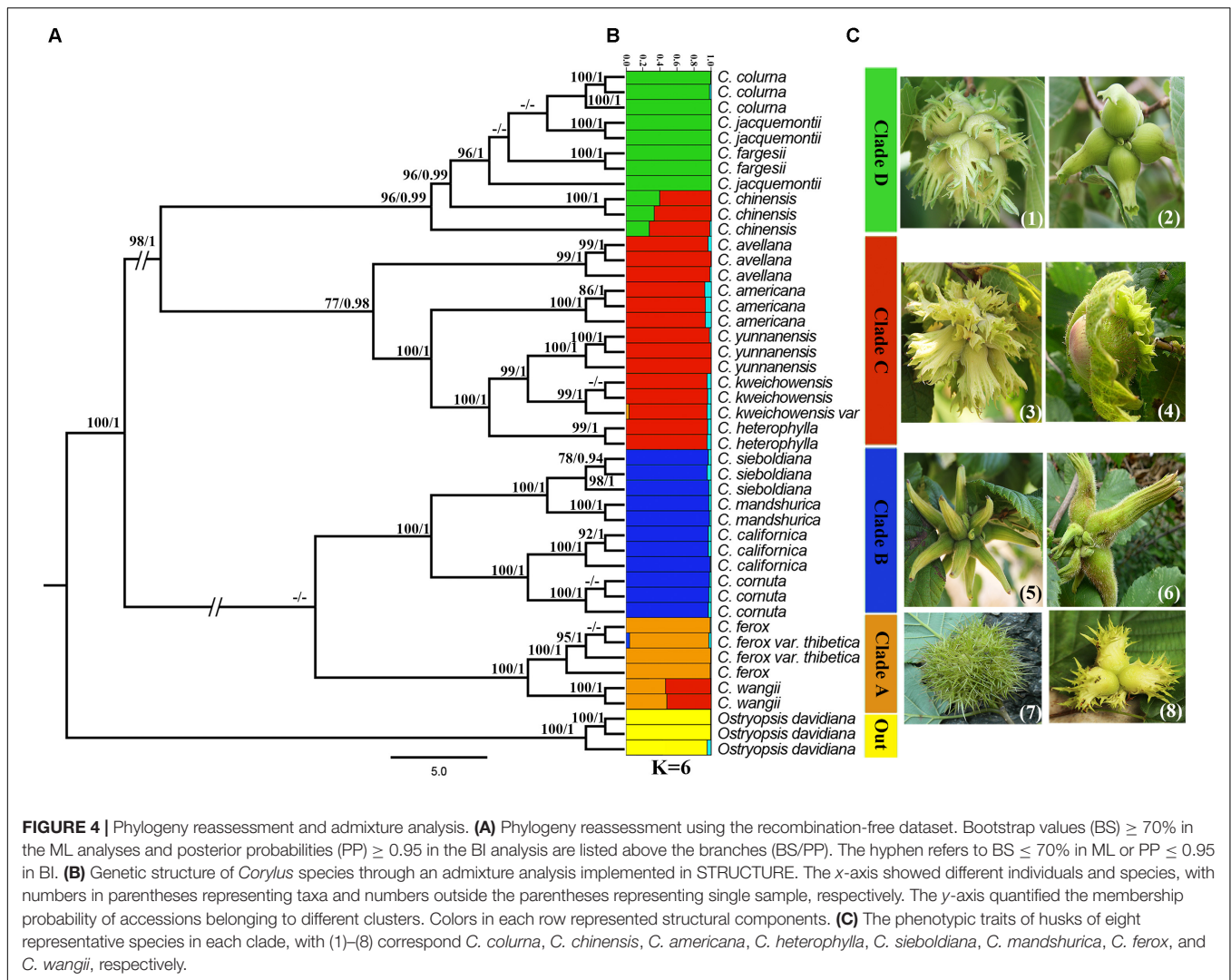


FIGURE 3 | Recombination events predicted using the RDP4 program. Three recombination events (Event 1–Event 3) supported by at least four methods are displayed. The putative recombinant segments associated with each recombination event are indicated in colors. The detailed positions of all recombination break points are also listed.

and BI using the recombination-free data (Figure 4A). All the nodes in the tree were highly supported with a posterior probability of >0.99 (Figure 5). The age of the most recent common ancestor (MRCA) of *Corylus* and *Ostryopsis* was

estimated by the BEAST analysis to be 51.54 Ma, with the 95% HPD ranging from 38.18 to 66.62 Ma. The MRCA of *Corylus* species began to occur in late Eocene (36.38 Ma, 95% HPD: 28.39–43.65 Ma), which was slightly earlier than the divergence



time of *Ostryopsis* (35.07 Ma, 95% HPD: 30.71–39.79 Ma). Within the genus *Corylus*, two major clades started to diverge in different directions at around 31.21 (95% HPD: 22.79–39.19 Ma) and 19.97 Ma (95% HPD: 12.81–28.47 Ma), respectively. In the early and middle Miocene (10.3–17.76 Ma), the rudiments of four modern clades have basically taken shape. Subsequently, rapid speciation in different clades occurred in the middle and late Miocene. The divergence between *C. wangii* and the section *Acanthochlamys* (*C. ferox* and *C. ferox* var. *thibetica*) in Clade A was estimated to be 17.76 Ma (95% HPD: 8.58–27.6 Ma). In Clade B, the intercontinental division between Asian species (*C. sieboldiana* and *C. mandshurica*) and American species (*C. cornuta* and *C. californica*) took place about 15.37 Ma (95% HPD: 9.9–21.38 Ma), while the split time within each subclade was identically around 10 Ma. The species differentiation in Clade C has experienced an incremental process, of which *C. avellana* and *C. americana* separated successively at about 16.13 and 13.52 Ma, while the *C. heterophylla* complex (*C. heterophylla*, *C. yunnanensis*, *C. kweichowensis*, and its variety) showed a rapid speciation (6.43–9.94 Ma). *C. chinensis* was the first to separate

from Clade D 10.3 Ma ago, whereas the divergence between *C. jacquemontii* and *C. colurna* was estimated to be 6.57 Ma.

Ancestral Area Reconstruction

The results of the ancestral area reconstruction were shown in **Figure 6**. BBM reconstruction indicated that southwestern China (D) was the most probable ancestral area for ancient *Corylus* species. Furthermore, results of the BBM analysis suggest that dispersal and vicariance played a substantial role in the biogeographic history of *Corylus*. Four long distance dispersal (LDD) events (nodes 1, 3, 4, and 5) from southwestern China to NA (route: D–C–B–A–G) and subsequent vicariance events (nodes 1, 3, and 5) were derived, forming four distinct lineages in Clade C: *C. americana*, *C. heterophylla*, *C. yunnanensis*, *C. kweichowensis*, and its variety *C. kweichowensis* var. *brevipes*. Almost the same route across the BLB (D–C–B–A–H), Clade B became an independent lineage through four dispersal (nodes 9, 10, 11, and 12) and two vicariance events (nodes 9 and 10). Furthermore, another two independent dispersal routes passing from southwestern China to the Himalayas (route: D–E; node 2)

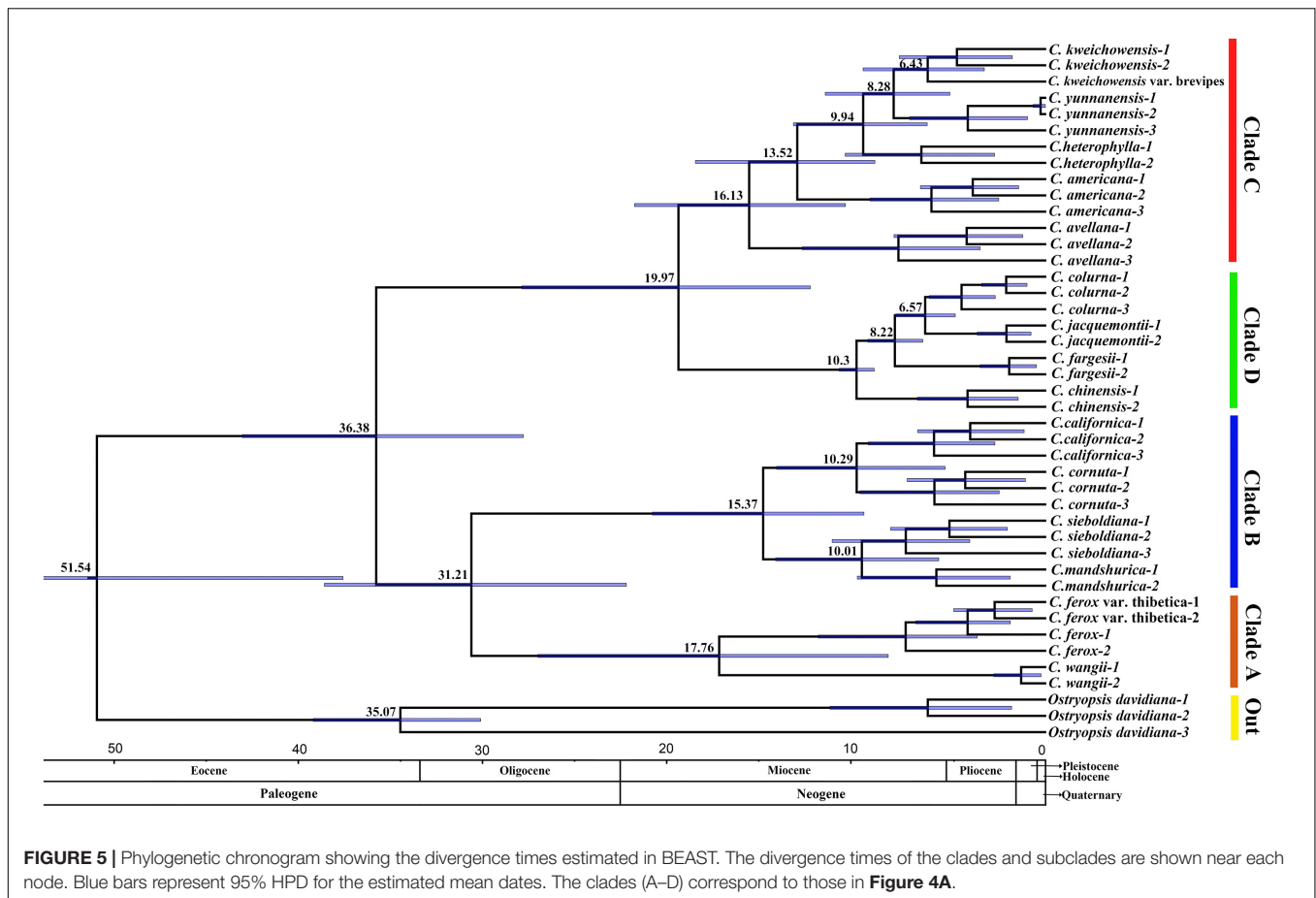


FIGURE 5 | Phylogenetic chronogram showing the divergence times estimated in BEAST. The divergence times of the clades and subclades are shown near each node. Blue bars represent 95% HPD for the estimated mean dates. The clades (A–D) correspond to those in **Figure 4A**.

and European-Mediterranean area (route: D–F; node 8) were also predicted. Along with subsequent vicariance events (nodes 2 and 8), *C. avellana* in Clade C and *C. jacquemontii* and *C. colurna* in Clade D became distinct species gradually. Besides, sympatric and parapatric speciation patterns with short distance dispersals were observed, including *C. wangii* and *C. ferox* in Clade A (nodes 13 and 14), and *C. chinensis* and *C. fargesii* in Clade D (nodes 6 and 7).

DISCUSSION

Speciation Mechanisms and Rapid Diversification

Results of the above analyses revealed strong recombination (**Table 2** and **Figure 3**) and hybridization signals (**Figure 4B**; **Supplementary Table S2**), as well as clear dispersal and vicariance events (**Figures 5, 6**), which have more or less influenced the speciation and diversification process of the genus *Corylus*. Phylogenies are always constructed assuming that nucleotide sequences replicate without recombining. However, it is very likely that recombination can severely bias population and phylogenetic analyses and finally lead to incorrect results (Martin et al., 2017; Kiil and Østerlund, 2018; Schierup and Hein, 2000). SNP nucleotide sequences of nuclear

genome have the characteristics of biparental inheritance, which potentially contain recombination events. Accordingly, inference of phylogeny using genome-wide SNPs can be severely distorted by recombination events either between sequences within the dataset or with an unobserved sequence. In this study, we discovered plenty of recombination events and simultaneously verified their influences on phylogeny. The fact indicates that all the four clades (A–D) have involved in recombination either as potential recombinants or parents (**Figures 2, 3**), of which almost all the taxa in Clade B and partial taxa in Clade D belong to recombinants while individuals from Clade A, Clade B, and Clade C are identified as recombinant parents. This universal phenomenon of interspecific recombination among *Corylus* species demonstrates the existence of hybridization, during which process the recombinants obtain the genetic components of both parents. The divergence of the two phylogenies based on concatenated and free-recombination dataset centers on the phylogenetic position of *C. avellana* which is classified into Clade C by the free-recombination dataset (**Figure 4A**) but ranked into Clade D by concatenated dataset (**Figure 2**). It turns out that the free-recombination phylogeny shows high consistency with previous morphological classification and ITS phylogeny (Erdogan and Mehlenbacher, 2000; Whitcher and Wen, 2001). Structure analysis also reveals extensive introgression existing in different clades,

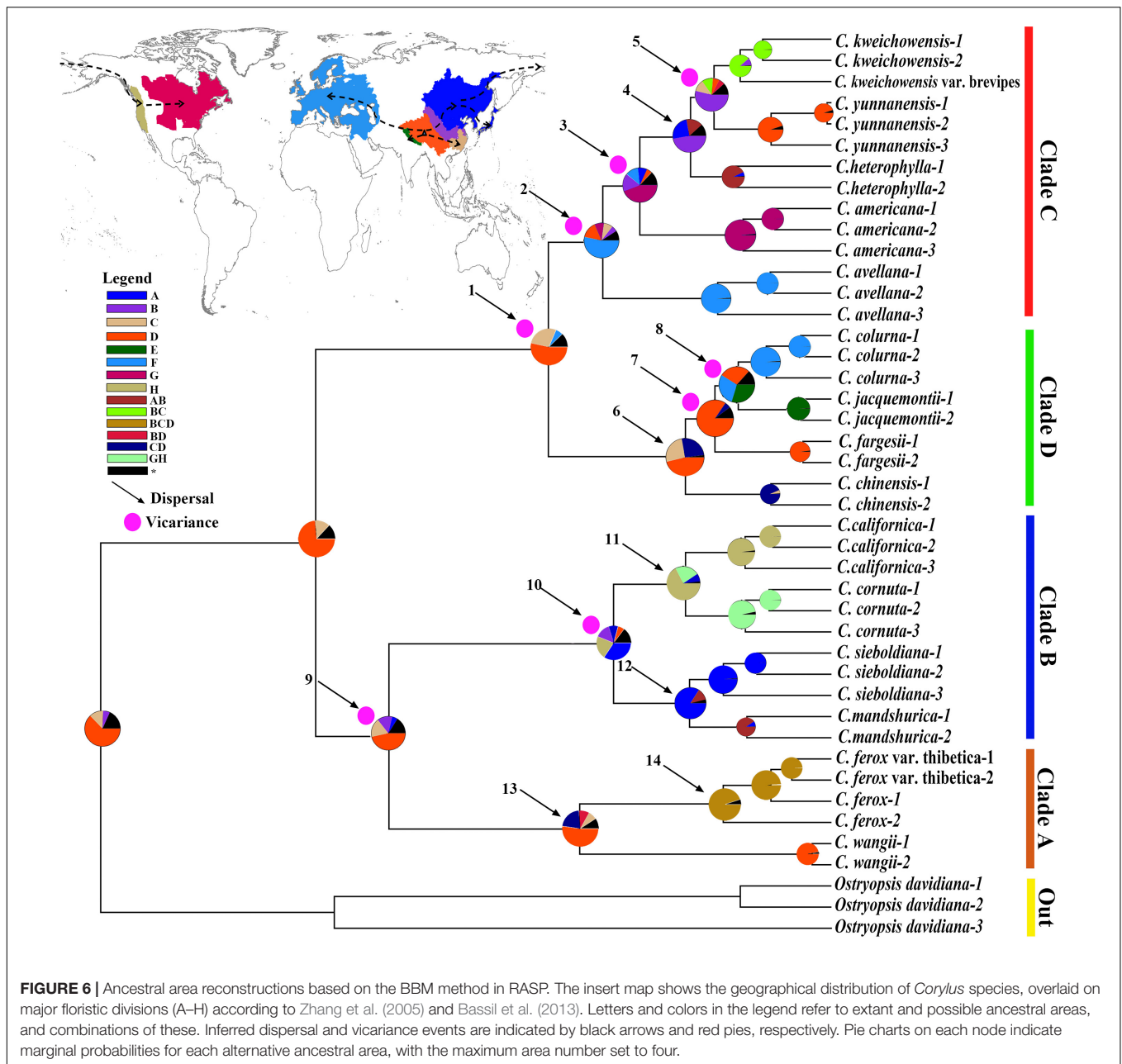


FIGURE 6 | Ancestral area reconstructions based on the BBM method in RASP. The insert map shows the geographical distribution of *Corylus* species, overlaid on major floristic divisions (A–H) according to Zhang et al. (2005) and Bassil et al. (2013). Letters and colors in the legend refer to extant and possible ancestral areas, and combinations of these. Inferred dispersal and vicariance events are indicated by black arrows and red pies, respectively. Pie charts on each node indicate marginal probabilities for each alternative ancestral area, with the maximum area number set to four.

especially evident between Clades C and A, and Clades C and D (Figure 4B). Similarly, hybridization signals are also predicted by NewHybrids detection (Supplementary Table S2). All the above analyses demonstrate that recombination and hybridization are important mechanisms for the speciation of *Corylus* species.

Dispersal–vicariance model is a significant mechanism in speciation, especially for the LDD. Divergence time estimation and ancestral area reconstruction indicate that the ancestors of *Corylus* originate from the southwestern China (D) in late Eocene (~36.38 Ma) (Figures 5, 6). Southwestern China, a region originated from the Qinghai–Tibetan Plateau (QTP) uplift, is finally shaped as one of the biodiversity hotspot in the north

temperate region and is regarded as the potential original center of many modern species (Li and Li, 1993; Wen et al., 2014; Deng et al., 2017). Thus, it is no wonder that *Corylus* has also originated from this diversity center. However, subsequent dispersal displays diverse patterns according to their dispersal routes. One of the most important routes involves the LDD from southwestern China to northeast Asia (D–C–B–A) and then across the BLB to NA (G/H), through which Asian bristle-husked shrubs (*C. sieboldiana* and *C. mandshurica*) in Clade B and leaf-husked shrub (*C. heterophylla*) in Clade C dispersed to NA at a very close time (~15.37 and 13.52 Ma). Rapid species diversification in the two clades occurred approximately between 6 and 16 Ma, during which time geological isolation

after dispersal may have played an important role. Besides, the ancestor of leafy husked shrubs is found to pass through West Asia to the European-Mediterranean area, which is slightly earlier (0.76 Ma) than the BLB dispersal, forming the famous European species *C. avellana* (Clade C). The modern species of Clade D are related with three speciation patterns: allopatric (*C. colurna* and *C. jacquemontii*), sympatric (*C. fargesii*), and parapatric (*C. chinensis*) speciation. The speciation time of Clade D falls into the late Miocene (6–10 Ma), which is fairly close to the divergence of several linkages in Clade C and is probably caused by the global geologic and climatic changes in East Asia during that time. Particularly, the rise of the Himalayas is probably the major cause to bring about the divergence between *C. colurna* and *C. jacquemontii*. Clade A differentiated firstly from *Corylus* at about 17.76 Ma, suggesting a more ancient origin than other clades. Notably, this clade involves extensive hybridization and recombination events, which are likely to be the genetic residual of ancient hybridization.

Evolutionary Relationships and Taxonomic Implications

Relationships among *Corylus* species have been studied previously either through phenotypic characters (Ferreira et al., 2010; Ciarmiello et al., 2014) or molecular markers (Erdogan and Mehlenbacher, 2000; Boccacci and Botta, 2010; Martins et al., 2014; Mohammadzadeh et al., 2014). However, none of them have reached definitive conclusions for the phylogenetic relationships and taxonomy of *Corylus* species partly due to incomplete taxa sampling by lacking some rare species such as *C. ferox*, *C. fargesii*, *C. wangii*, and *C. chinensis*, and partly because of the low resolution of molecular markers including SSR marker, *matK* gene, and ITS regions. In our study, detailed sampling was conducted by collecting almost all the *Corylus* species generally accepted in the current, and integrated analyses were performed by combining phylogenetics and structure inference. Although ML and BI phylogenies revealed identical classification results by dividing the ingroup into four clades (Figure 2), it is confirmed that the phylogenetic topology has been influenced by recombination events (Figure 3), which finally resulted in the misjudgment on the phylogenetic position of *C. avellana* (Figure 2). However, a robust phylogeny is successfully reconstructed based on the integrated inference from structure analysis and recombination tests (Figure 4A). That is, four distinct clades (A–D) are identified by transferring *C. avellana* from Clade D to Clade C. In the following, we will discuss each of these clades, respectively.

Clade A

Clade A is comprised of three ancient species (variety): *C. wangii*, *C. ferox*, and its variety *C. ferox* var. *thibetica*. *C. ferox* was once suggested to be the basalmost extant taxon of the genus based on morphological traits and ITS phylogeny (Li and Cheng, 1979; Whitcher and Wen, 2001). It is very distinct from other *Corylus* species in fruit involucre in that it has bur-like spiny

husks that resemble those of chestnuts (Figure 4C). Despite that *C. ferox* and *C. ferox* var. *thibetica* are not well separated by molecular phylogeny in this study, potential phenotypic differentiation can still provide valid evidence to distinguish them. The visible difference lies in their leaf size and shape, with leaves of the variety being more spacious than native species (Figures 1L,M). *C. wangii* has been seldomly studied by researchers due to its limited geographical distribution. However, it is definitely a unique species that segregates from other *Corylus* species but displayed a close affiliation with *C. ferox*. Integrated analyses of structure analysis (Figure 4B) and hybridization detection (Supplementary Table S2) indicate that *C. wangii* is probably originated from ancient hybridization between Clades A and C. Molecular dating analysis supports the earliest differentiation of Clade A (Figure 5), revealing an ancient origin as described above. However, the resulted phylogeny does not reflect the monophyletic or basal position of this clade but assign it as the sister to Clade B. The discordance may probably be caused by incomplete sampling for *Corylus* species or the low resolution of markers in previous studies.

Clade B

Clade B is the most robust paraphyletic group that consists of two North American species *C. californica* and *C. cornuta*, and two East Asian species *C. sieboldiana* and *C. mandshurica*. This clade in our results is in agreement with those of Erdogan and Mehlenbacher (2000), Whitcher and Wen (2001), and Bassil et al. (2013). The remarkable features of these species are their tubular and beaked husks with very loosely attached bristles (Figures 1Q,X, 4C). Although these species exhibit disjunctive distribution between East Asia and NA, multivariate data from interspecific hybridization relations, phenotypic characters, and molecular markers reveal high similarities and close affinity among these species. Distribution of *C. mandshurica* displays apparent dispersal from southwest to northeast in China, while *C. sieboldiana* mainly distributes in the adjacent regions such as Korean peninsula and Japanese archipelago. Thompson et al. (1996) proposed *C. mandshurica* was synonyms or variety of *C. sieboldiana*, and Chin et al. (2004) noted that the two belonged to different populations of the same species. In view of their separate distribution areas and the inferred dispersal route, it is obvious that *C. mandshurica* is the original species, while *C. sieboldiana* is a derivative. Therefore, we do not support the designation that *C. mandshurica* was the variety of *C. sieboldiana*. On the contrary, we suggest *C. sieboldiana* to be a variety of *C. mandshurica* or a distinct species. *C. californica* and *C. cornuta* distribute in two different regions of NA: western NA (H) and eastern NA (G), respectively. The taxonomic status of *C. californica* has long been controversial. It was once viewed as a distinct species by some researchers (Krüssmann, 1976; Erdogan and Mehlenbacher, 2000), but as a botanical variety (Everett, 1982; Thompson et al., 1996; Huxley and Griffiths, 1999) or a subspecies (Furlow, 1997) of *C. cornuta* by others. In the present study, *C. californica* and *C. cornuta* formed their separate group in the phylogeny (Figure 4A), suggesting the genetic divergence between them. Notably, ancestral area reconstruction reveals a

dispersal route from western NA (H) to eastern NA (G), which probably means that the occurrence of *C. cornuta* was later than *californica*. Moreover, the easy cross between the two may support the botanical variety designation. Similarly, we would put forward a different opinion that *C. californica* is the product of LDD crossing BLB of Asian *C. mandshurica* and *C. sieboldiana*, while *C. cornuta* is a geographical variety of *C. californica* or have become a distinct species. The deep-rooted relationships in this clade remain to be further investigated through population genetics.

Clade C

Clade C is formed by six leafy-husked shrubs that disjunctively distribute among East Asia, Europe, and NA, of which, four Chinese species (*C. heterophylla*, *C. yunnanensis*, *C. kweichowensis*, and its variety *C. kweichowensis* var. *brevipes*.) constitute the *C. heterophylla* complex, while the other two species (*C. americana* and *C. avellana*) form their separate subclade. It is interesting that these species group together although they are geographically isolated. The potential relationships among them can be described using the morphological similarity especially for husks and nuts. The typical husk characters of these species are leafy-shape, with deep incisions in the margin and no constriction at the tip (Figures 1P, 4C). Furthermore, it is verified that the three species *C. americana*, *C. avellana*, and *C. heterophylla* can hybridize easily with each other under natural conditions, with the hybrid variety "*C. heterophylla* × *C. avellana*" as a typical case. Divergence time estimation and ancestral area reconstruction indicate that *C. avellana* and *C. americana* successively spread to Europe and NA via two different dispersal routes in early and middle Miocene, respectively. Species differentiation in the *C. heterophylla* complex is found to occur recently, which may lead to the difficulty in species identification. Once, *C. kweichowensis* and *C. yunnanensis* were both viewed as botanical varieties of *C. heterophylla* by some researchers (Yu, 1979; Thompson et al., 1996) and as distinct species by others (Liang and Zhang, 1988; Qi, 1996; Ma et al., 2014). Besides, *C. kweichowensis* var. *brevipes*, a variety with typically brachypodous characteristics was also identified (Liang and Zhang, 1988). In our study, the phylogenies support the close affinity among *C. heterophylla*, *C. kweichowensis*, and *C. yunnanensis* (Figure 4A). *C. heterophylla* and *C. yunnanensis* that locate at two terminals of the dispersal route differentiate earlier than *C. kweichowensis* in the middle zone. We infer that the significant differences in ecological conditions between south and north of China as well as the frequent gene flow in the middle zone have affected the species differentiation in this complex. Remarkably, a vicariance event (node 5 in Figure 6) occurred in the late Miocene between *C. kweichowensis* and *C. yunnanensis*, providing more support for the designation of distinct species.

Clade D

Clade D is a well-resolved phylogenetic group composed by four tree species: *C. chinensis*, *C. colurna*, *C. jacquemontii*, and *C. fargesii*. The clustering group of *C. chinensis*, *C. colurna*, and *C. jacquemontii* was previously demonstrated by nuclear

microsatellite-based clustering (Bassil et al., 2013), strict consensus tree of *matK* gene, and ITS regions (Erdogan and Mehlenbacher, 2000; Whitcher and Wen, 2001). Although there are few reports about the molecular taxonomy of *C. fargesii* in previous studies, both the phylogenetic and structure analysis classify this species into the Clade D. Therefore, we support Clade D as a reliable clustering group. The common phenotypic characters of these four species are their big single stems and juicy-flesh husks covered with glandular hairs (Figures 1C–F, 4C). Both *C. chinensis* and *C. fargesii* own the constricted tubular husks that prevent the nuts falling even after maturity (Figures 1R,S). *C. fargesii*, also called the paperbark tree hazel, is native to China and morphologically distinct from other tree species, with its bark exfoliating like *Betula* species (Figure 1Y). *C. colurna* from the Mediterranean regions is recognized as the carbon copy of *C. jacquemontii* that distributes in the Himalayas by presenting not only extremely similar growth habit, but also analogous husk and nut morphology. We infer that the divergence between the two is caused by the rise of the Himalayas.

AUTHOR CONTRIBUTIONS

T-TZ and G-XW conceived and designed the experiments. ZY, T-TZ, G-XW, Q-HM, and L-SL participated in the collection of study materials. ZY and T-TZ participated in the DNA extraction and data analysis. ZY wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 31500555), and the Special Fund for Basic Scientific Research Business of Central Public Research Institutes (Grant Nos. CAFYBB2016QB003 and CAFYBB2017ZA004-9).

ACKNOWLEDGMENTS

The authors sincerely thank Dr. Zhaoshan Wang of the Chinese Academy of Forestry and Shanghai OE Biomedical Technology Co., Ltd. for assistance with this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01386/full#supplementary-material>

FIGURE S1 | The log probability of the ΔK -value for 10 replicated STRUCTURE runs given each number of clusters *K* based on genome-wide SNPs.

TABLE S1 | Summary of high-throughput sequencing.

TABLE S2 | The estimated posterior probabilities generated with NewHybrids that each individual belongs to each of the different genotype frequency categories.

REFERENCES

- Anderson, E. C., and Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genet data. *Genetics* 160, 1217–1299.
- Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., et al. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 113, 8320–8325. doi: 10.1073/pnas.1600405113
- Bassil, N. V., Boccacci, P., Botta, R., Postman, J., and Mehlenbacher, S. (2013). Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars. *Genet. Resour. Crop Evol.* 60, 543–568. doi: 10.1007/s10722-012-9857-z
- Bobrov, E. G. (1936). “Coryleae,” in *Flora of the U.S.S.R.* 5, ed. V. L. Komarov (Leningrad: Izdatel'stvo Akademii Nauk SSSR), 253–268.
- Boccacci, P., and Botta, R. (2010). Microsatellite variability and genetic structure in hazelnut (*Corylus avellana* L.) cultivars from different growing regions. *Sci Hortic.* 124, 128–133. doi: 10.1016/j.scienta.2009.12.015
- Boni, M. F., Posada, D., and Feldman, M. W. (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics* 176, 1035–1047. doi: 10.1534/genetics.106.068874
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C. H., Xie, D., et al. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Comput. Biol.* 10:e1003537. doi: 10.1371/journal.pcbi.1003537
- Boufford, D. E., and Spongberg, S. A. (1983). Eastern asian-eastern north American phytogeographical relationships—a history from the time of linnaeus to the twentieth century. *Ann. Missouri Bot. Garden* 70, 423–439. doi: 10.2307/2992081
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci de novo from short-read sequences. *G3* 1, 171–182. doi: 10.1534/g3.111.000240
- Chen, C., Qi, Z. C., Xu, X. H., Comes, H. P., Koch, M. A., Jin, X. J., et al. (2014). Understanding the formation of Mediterranean–African–Asian disjunctions: evidence for miocene climate-driven vicariance and recent long-distance dispersal in the Tertiary relict *Smilax aspera* (Smilacaceae). *New Phytol.* 204, 243–255. doi: 10.1111/nph.12910
- Chin, S. C., Chang, G.-S., and Qin, H. N. (2004). A multivariate morphometric study on *Corylus sieboldiana* complex (Betulaceae) in China, Korea, and Japan. *Acta Phytotaxon. Sin.* 42, 222–235.
- Chen, Z. D., Manchester, S. R., and Sun, H. Y. (1999). Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology, and paleobotany. *Am. J. Bot.* 86, 1168–1181. doi: 10.2307/2656981
- Ciarmiello, L. F., Mazzeo, M. F., Minasi, P., Peluso, A., De, Luca, A., Piccirillo, P., et al. (2014). Analysis of different European hazelnut (*Corylus avellana* L.) cultivars: authentication, phenotypic features, and phenolic profiles. *J. Agric. Food Chem.* 62, 6236–6246. doi: 10.1021/jf5018324
- Crane, P. R. (1989). “Early fossil history and evolution of the Betulaceae,” in *Evolution, Systematics, and Fossil History of the Hamamelidae*, Vol. 2, eds P. R. Crane and S. Blackmore (Oxford: Clarendon Press), 87–116.
- Deng, T., Zhang, J. W., Meng, Y., Volis, S., Sun, H., and Nie, Z. L. (2017). Role of the qinghai-tibetan plateau uplift in the Northern Hemisphere disjunction: evidence from two herbaceous genera of Rubiaceae. *Sci. Rep.* 7:13411. doi: 10.1038/s41598-017-13543-5
- Donoghue, M. J., and Smith, S. A. (2004). Patterns in the assembly of temperate forests around the northern hemisphere. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 359, 1633–1644. doi: 10.1098/rstb.2004.1538
- Earl, D. A. (2012). Structure harvester: a website and program for visualizing Structure output and implementing the Evanno method. *Conserv Genet Resour.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Editorial Board of the Flora of China of Chinese Academy of Sciences (2013). *Flora of China*, Vol. 21. Beijing: Science Press.
- Erdogan, V., and Mehlenbacher, S. A. (2000). Phylogenetic relationships of *Corylus* species (Betulaceae) based on nuclear ribosomal DNA ITS region and chloroplast matK gene sequences. *Syst. Bot.* 25, 727–737. doi: 10.2307/2666730
- Everett, T. H. (1982). *The New York Botanical Garden Illustrated Encyclopedia of Horticulture*, Vol. 10. Milton Park: Taylor & Francis.
- Ferreira, J. J., Garcia, G. C., Tous, J., and Rovira, M. (2010). Genetic diversity revealed by morphological traits and ISSR markers in hazelnut germplasm from northern Spain. *Plant Breed.* 129, 435–441. doi: 10.1111/j.1439-0523.2009.01702.x
- Fu, X., Dou, J., Mao, J., Su, H., Jiao, W., Zhang, L., et al. (2013). RADtyping: an integrated package for accurate de novo codominant and dominant RAD genotyping in mapping populations. *PLoS One* 8:e79960. doi: 10.1371/journal.pone.0079960
- Furrow, J. J. (1997). “Betulaceae Gray, Birch family,” in *Flora of North America Magnoliophyta: Magnoliidae and Hamamelidae*, Vol. 3, ed. N. R. Morin (New York, NY: Oxford Univ. Press), 507–538.
- Gibbs, M. J., Armstrong, J. S., and Gibbs, A. J. (2000). Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics* 16, 573–582. doi: 10.1093/bioinformatics/16.7.573
- Govaerts, R., and Frodin, D. G. (1998). *World Checklist and Bibliography of Fagales (Betulaceae, Corylaceae, Fagaceae and Ticoendraceae)*. Richmond: The Royal Botanic Gardens Kew.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Holstein, N., el Tamer, S., and Weigend, M. (2018). The nutty world of hazel names—a critical taxonomic checklist of the genus *Corylus* (Betulaceae). *Eur. J. Taxonomy* 409, 1–45. doi: 10.5852/ejt.2018.409
- Holt, R. D., and Gomulkiewicz, R. (1997). How does immigration influence local adaptation? A reexamination of a familiar paradigm. *Am. Nat.* 149, 563–572. doi: 10.1086/286005
- Huxley, A. J., and Griffiths, M. (1999). *New Royal Horticultural Society Dictionary of Gardening*. New York, NY: Grove's Dictionaries Inc.
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von, Haeseler A., and Jermin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587. doi: 10.1038/nmeth.4285
- Kelchner, S. A., and Bamboo Phylogeny Group (2013). Higher level phylogenetic relationships within the bamboos (Poaceae: Bambusoideae) based on five plastid markers. *Mol. Phylogenet. Evol.* 67, 404–413. doi: 10.1016/j.ympev.2013.02.005
- Kiil, K., and Østerlund, M. (2018). CleanRecomb, a quick tool for recombination detection in SNP based cluster analysis. *bioRxiv* [Preprint]. doi: 10.1101/317131
- Krüssmann, G. (1976). *Manual of Cultivated Broad-leaved Trees and Shrubs*. Portland: Timber Press
- Li, P. C., and Cheng, S. X. (1979). “Betulaceae,” in *Flora Republicae Popularis Sinicae*, Vol. 21, eds K.-Z. Kuang and P.-C. Li (Henderson, NV: Science Press), 44–137.
- Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., et al. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336
- Li, X. W., and Li, J. (1993). A preliminary foristic study on the seed plants from the region of Hengduan Mountain. *Acta Bot. Yunnan* 15, 217–231.
- Liang, W. J., and Zhang, Y. M. (1988). “Investigation and study of filbert resources in China,” in *Proceedings of the International Symposium on Horticultural Germplasm, Cultivated and Wild*, Beijing, 5–9.
- Livingstone, K., and Rieseberg, L. (2004). Chromosomal evolution and speciation: a recombination-based approach. *New Phytol.* 161, 107–112. doi: 10.1046/j.1469-8137.2003.00942.x
- Ma, Q. H., Huo, H. L., Chen, X., Zhao, T. T., Liang, W. J., and Wang, G. X. (2014). Study on the taxonomy, distribution, development and utilization of *Corylus kweichowensis* Hu. *J. Plant Genet. Resour.* 15, 1223–1231.
- Mao, K., Hao, G., Liu, J., Adams, R. P., and Milne, R. I. (2010). Diversification and biogeography of Juniperus (Cupressaceae): variable diversification rates and multiple intercontinental dispersals. *New Phytol.* 188, 254–272. doi: 10.1111/j.1469-8137.2010.03351.x
- Martin, D. P., Murrell, B., Khoosal, A., and Muhire, B. (2017). *Detecting and Analyzing Genetic Recombination Using RDP4*. New York, NY: Humana Press, 433–460.
- Martin, D. P., Murrell, B., Golden, M., Khoosal, A., and Muhire, B. (2015). RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 1:vev003. doi: 10.1093/ve/vev003
- Martin, D. P., Posada, D., Crandall, K. A., and Williamson, C. (2005). A modified bootscan algorithm for automated identification of recombinant sequences and

- recombination breakpoints. *AIDS Res. Hum. Retroviruses* 21, 98–102. doi: 10.1089/aid.2005.21.98
- Martin, D. P., and Rybicki, E. (2000). RDP: detection of recombination amongst aligned sequences. *Bioinformatics* 16, 562–563. doi: 10.1093/bioinformatics/16.6.562
- Martins, S., Simões, F., Matos, J., Silva, A. P., and Carnide, V. (2014). Genetic relationship among wild, landraces and cultivars of hazelnut (*Corylus avellana*) from Portugal revealed through ISSR and AFLP markers. *Plant Syst. Evol.* 300, 1035–1046. doi: 10.1007/s00606-013-0942-3
- Migliore, J., Baumel, A., Juin, M., and Médail, F. (2012). From Mediterranean shores to central Saharan mountains: key phylogeographical insights from the genus *Myrtus*. *J. Biogeogr.* 39, 942–956. doi: 10.1111/j.1365-2699.2011.02646.x
- Minh, B. Q., Nguyen, M. A. T., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195. doi: 10.1093/molbev/mst024
- Mohammadzadeh, M., Fattahi, R., Zamani, Z., and Khadivi-Khub, A. (2014). Genetic identity and relationships of hazelnut (*Corylus avellana* L.) landraces as revealed by morphological characteristics and molecular markers. *Sci. Hortic.* 167, 17–26. doi: 10.1016/j.scienta.2013.12.025
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2014). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Ohwi, J. (1965). *Flora of Japan [in English]*. Washington, DC: Smithsonian Institution.
- Padidam, M., Sawyer, S., and Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* 265, 218–225. doi: 10.1006/viro.1999.0056
- Peakall, R., and Smouse, P. E. (2012). GenALEX 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28, 2537–2539. doi: 10.1111/j.1471-8286.2005.01155.x
- Posada, D., and Crandall, K. A. (2001). Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13757–13762. doi: 10.1073/pnas.241370698
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Qi, J. Z. (1996). A study on classification of *Corylus kweichowensis*. *J. Nanjing For. Univ.* 20, 71–74.
- Rambaut, A. (2012). *FigTree v1.4*. Edinburgh: University of Edinburgh.
- Rambaut, A., and Drummond, A. J. (2014). *TreeAnnotator v2.1.2*. Edinburgh: University of Edinburgh.
- Ronquist, F., Teslenko, M., van der Mark P, Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Schierup, M. H., and Hein, J. (2000). Consequences of recombination on traditional phylogenetic analysis. *Genetics* 156, 879–891.
- Smith, J. M. (1992). Analyzing the mosaic structure of genes. *J. Mol. Evol.* 34, 126–129. doi: 10.1007/BF00182389
- Sobel, J. M., and Streisfeld, M. A. (2015). Strong premating reproductive isolation drives incipient speciation in *Mimulus aurantiacus*. *Evolution* 69, 447–461. doi: 10.1111/evo.12589
- Takhtajan, A. L. (1982). *Fossil flowering plants of the USSR: Ulmaceae-Betulaceae*, Vol. 2. Leningrad: Nauka.
- Thompson, M. M., Lagerstedt, H. B., and Mehlenbacher, S. A. (1996). “Hazelnuts,” in *Fruit Breeding—Volume III. Nuts*, eds J. Janick and N. Moore (New York: John Wiley & Sons), 125–184.
- Tiffney, B. H., and Manchester, S. R. (2001). The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the northern hemisphere tertiary. *Int. J. Plant Sci.* 162, S3–S17. doi: 10.1086/323880
- Tutin, T. G., and Walters, S. M. (1993). “Corylaceae,” in *Flora Europaea*, Vol. 1, eds T. G. Tutin, N. A. Burges, A. O. Chater, J. R. Edmondson, V. H. Heywood, D. M. Moore, et al. (Cambridge: Cambridge University Press), 70–71.
- Wang, N., Thomson, M., Bodles, W. J., Crawford, R. M., Hunt, H. V., Featherstone, A. W., et al. (2013). Genome sequence of dwarf birch (*Betula nana*) and cross-species RAD markers. *Mol. Ecol.* 22, 3098–3111. doi: 10.1111/mec.12131
- Wang, S., Meyer, E., McKay, J. K., and Matz, M. V. (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat. Methods* 9, 808–810. doi: 10.1038/nmeth.2023
- Wen, J. (1999). Evolution of eastern Asian and eastern North American disjunct distributions of flowering plants. *Annu. Rev. Ecol. Syst.* 30, 421–455. doi: 10.1146/annurev.ecolsys.30.1.421
- Wen, J., and Ickert-bond, S. M. (2009). Evolution of the Madrean–Tethyan disjunctions and the North and South American amphitropical disjunctions in plants. *J. Syst. Evol.* 47, 331–348. doi: 10.1111/j.1759-6831.2009.00054.x
- Wen, J., Zhang, J., Nie, Z. L., Zhong, Y., and Sun, H. (2014). Evolutionary diversifications of plants on the Qinghai-Tibetan Plateau. *Front. Genet.* 5:4. doi: 10.3389/fgene.2014.00004
- Whitcher, I. N., and Wen, J. (2001). Phylogeny and biogeography of *Corylus* (Betulaceae): inferences from ITS sequences. *Syst. Bot.* 26, 283–298. doi: 10.1043/0363-6445-26.2.283
- Wolf, D. E., Takebayashi, N., and Riesenber, L. H. (2001). Predicting the risk of extinction through hybridization. *Conserv. Biol.* 15, 1039–1053. doi: 10.1046/j.1523-1739.2001.0150041039.x
- Wolfe, J. A., and Wehr, W. (1987). *Middle Eocene Dicotyledonous Plants From Republic, Northeastern Washington*. Washington DC: U.S. Government Printing Office.
- Xiang, J. Y., Wen, J., and Peng, H. (2015). Evolution of the eastern Asian–North American biogeographic disjunctions in ferns and lycophytes. *J. Syst. Evol.* 2015, 2–32. doi: 10.1111/jse.12141
- Xiang, Q., and Soltis, D. E. (2001). Dispersal–vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the northern hemisphere. *Int. J. Plant Sci.* 162, 29–39. doi: 10.1086/323332
- Xiang, Q. Y., Zhang, W. H., Ricklefs, R. E., Qian, H., Chen, Z. D., Wen, J., et al. (2004). Regional differences in rates of plant speciation and molecular evolution: a comparison between eastern Asia and eastern North America. *Evolution* 58, 2175–2184. doi: 10.1554/03-712
- Yu, D. J. (1979). *Deciduous Fruit Tree Taxonomy*. Shanghai: Shanghai Science and Technology Press.
- Yu, Y., Harris, A. J., Blair, C., and He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Mol. Phylogenet. Evol.* 87, 46–49. doi: 10.1016/j.ympev.2015.03.008
- Zhang, J. B. (2014). *Biogeographical Studies of Fagales*. Doctoral thesis, Chinese Academic of Sciences, Beijing.
- Zhang, Y. H., Liu, L., Liang, W. J., and Zhang, Y. M. (2005). *Chinese Fruit Trees Record—Chestnut and Hazelnut*. Beijing: China Forestry Publishing House, 202–214.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Yang, Zhao, Ma, Liang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.