Phylogenomics Reveals an Ancient Hybrid Origin of the Persian Walnut

Bo-Wen Zhang,[†,1] Lin-Lin Xu,[†,1] Nan Li,[†,1] Peng-Cheng Yan,[2] Xin-Hua Jiang,[1] Keith E. Woeste,[3] Kui Lin,[*,1] Susanne S. Renner,[*,4] Da-Yong Zhang,[*,1] and Wei-Ning Bai[*,1]

[1]State Key Laboratory of Earth Surface Processes and Resource Ecology and Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China

[2]Beijing Key Laboratory of Cloud Computing Key Technology and Application, Beijing Computing Center, Beijing, China

[3]USDA Forest Service Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, West Lafayette, IN

[4]Department of Biology, Systematic Botany and Mycology, University of Munich (LMU), Munich, Germany

[†]These authors contributed equally to this work.

***Corresponding authors:** E-mails: linkui@bnu.edu.cn; renner@lmu.de; zhangdy@bnu.edu.cn; baiwn@bnu.edu.cn.

Associate editor: Claudia Russo

Abstract

Persian walnut (*Juglans regia*) is cultivated worldwide for its high-quality wood and nuts, but its origin has remained mysterious because in phylogenies it occupies an unresolved position between American black walnuts and Asian butternuts. Equally unclear is the origin of the only American butternut, *J. cinerea*. We resequenced the whole genome of 80 individuals from 19 of the 22 species of *Juglans* and assembled the genome of its relatives *Pterocarya stenoptera* and *Platycarya strobilacea*. Using phylogenetic-network analysis of single-copy nuclear genes, genome-wide site pattern probabilities, and Approximate Bayesian Computation, we discovered that *J. regia* (and its landrace *J. sigillata*) arose as a hybrid between the American and the Asian lineages and that *J. cinerea* resulted from massive introgression from an immigrating Asian butternut into the genome of an American black walnut. Approximate Bayesian Computation modeling placed the hybrid origin in the late Pliocene, ~3.45 My, with both parental lineages since having gone extinct in Europe.

Key words: **Approximate Bayesian Computation, hybridization, Juglans, phylogenetic networks, phylogeny.**

Introduction

Persian walnut (*Juglans regia* L.) is a nut crop of considerable economic importance. According to FAO statistics (http://www.fao.org/, accessed December 2018), China leads world production with 350,000 tons in 1997, followed by California, Turkey, and Iran. The origin and evolutionary history of the Persian walnut, however, are not understood, complicating the development of strategies to conserve germplasm. The genus *Juglans* (walnuts and butternuts) consists of ~22 species distributed in the Americas, southeastern Europe, and eastern Asia (supplementary fig. S1, Supplementary Material online, shows species distributions). All are wind pollinated and diploid (with $2n = 32$ chromosomes), and many are known to hybridize in the wild and in cultivation. Taxonomic studies commonly accept three or four sections based on fruit morphology and foliage architecture (Manning 1978; Manchester 1987). Section *Juglans* (including the younger synonym *Dioscaryon*) consists of the Persian walnut, native to Eurasia, and *J. sigillata* (known as Iron walnut), an ecotype maintained as a landrace in southwestern China and hybridizing naturally with Persian walnut (Wang et al. 2015; Zhao et al. 2018). Section *Rhysocaryon* (black walnuts) includes 16 species from North America, Central America, and South America (Stone et al. 2009), and section *Cardiocaryon* (butternuts or white walnuts) includes three species from eastern Asia (Lu et al. 1999) and one, *J. cinerea*, from eastern North America.

Phylogenies based on RFLPs and plastid and nuclear loci have supported major clades of North American, Asian, and Persian walnuts but have been unable to resolve their relationships to each other (Fjellstrom and Parfitt 1995; Stanford et al. 2000; Manos and Stone 2001; Aradhya et al. 2007; Stone et al. 2009; Dong et al. 2017). Equally unclear is the phylogenetic position of the American *J. cinerea*, which shifts from being a member of the Asian butternut clade in nuclear trees to becoming a member of the North American black walnut clade in plastid trees (Aradhya et al. 2007; Dong et al. 2017). In pilot analyses using quartet frequencies on 2901 single-copy nuclear genes from 19 species, we were able to exclude incomplete lineage sorting as the cause of the phylogenetic uncertainty (supplementary section S1 and tables S1 and S2, Supplementary Material online), leading us to speculate that ancient hybridization might be involved in the origin of the Persian walnut and the American butternut. Hybridization has played a central role in the origin of many crops, including apple (Cornille et al. 2012), banana (Christelová et al. 2011), *Citrus* (Wu et al. 2018), sweet potato

Article

(Munoz-Rodriguez et al. 2018), and wheat (El Baidouri et al. 2017).

To test our novel hypothesis of ancient hybridization in the walnut genus, we here use whole-genome-sequencing data from 80 individuals representing 19 of the 22 species of *Juglans*. Specifically, we asked whether the ancestral Persian walnut results from hybridization in the deep past, involving Asian and American ancestors. We apply a battery of genome-wide methods for hybridization detection and Approximate Bayesian Computation (ABC) to test speciation models and to infer the time of origin of the Persian walnut. Lastly, we characterize the genetic composition of the genomes of the Persian walnut, the Iron walnut, and the American butternut *J. cinerea*, by using population-genetic parameters, admixture analyses, and phylogenetic inference.

Results

Genome Sequencing and Variant Calling

Besides the 80 individuals of *Juglans*, we sequenced and assembled two outgroup species, *Pterocarya stenoptera* and *Platycarya strobilacea*, de novo to serve as reference genomes (Materials and Methods and supplementary table S2 and fig. S1, Supplementary Material online). The assembly of the *P. stenoptera* genome had a total length of 671 Mb (N50 = 1.28 Mb), with 548 scaffolds (containing 520 Mb) of length >100 kb. The assembly of the *Pl. strobilacea* genome comprised a total length of 678 Mb (N50 = 0.99 Mb) with 926 scaffolds (containing 649 Mb) of length >100 kb (supplementary section S1 and table S3, Supplementary Material online). Because *P. stenoptera* is equally distant from all *Juglans* taxa, reads of all 80 *Juglans* individuals were mapped to the *P. stenoptera* genome, covering an average of about 70.6% of that genome. To produce a high-quality genome-wide single-nucleotide polymorphism (SNP) data set for comparisons across *Juglans*, we focused on biallelic SNPs covered by all species from scaffolds with a length >100 kb. Ultimately, 19,795 SNPs were obtained after the total was thinned based on a 5-kb distance minimum, as linkage disequilibrium decays within this distance (supplementary fig. S2, Supplementary Material online), and a minor allele frequency (MAF) >0.05. In addition, we selected 2,901 single-copy genes to reconstruct a nuclear phylogeny (Materials and Methods; supplementary section S1, Supplementary Material online).

Population Structure within the Genus *Juglans*

In a STRUCTURE analysis of the 19,795 SNPs, both values of the log-likelihood of the data, ln Pr(*K*), and delta *K* (supplementary fig. S3, Supplementary Material online), indicated that the optimal value for *K* (i.e., the number of clusters) was 3 (fig. 1*a*). At $K = 2$, all black walnuts (section *Rhysocaryon*) clustered in one group and all butternuts (section *Cardiocaryon*) in another, whereas *J. regia* and *J. sigillata* (section *Juglans*) appeared to be an admixed group (fig. 1*a*), with 67% ancestry attributed to *Rhysocaryon* and 33% to *Cardiocaryon*. At $K = 3$, samples of section *Juglans* formed a distinct group. With *J. mandshurica* selected to represent *Cardiocaryon* and *J. nigra* for *Rhysocaryon*, the genetic ancestry of *J. regia* and *J. sigillata* can be ascertained in an admixture analysis of these four species by setting $K = 2$ with USEPOPINFO = 1 for parental *J. mandshurica* and *J. nigra* in STRUCTURE (based on 26,995 SNPs that meet the criteria of a 5-kb distance minimum and MAF > 0.05 for these four species). In this four-species analysis, ~51% of the genetic composition of *J. regia* or *J. sigillata* came from *J. nigra* and 49% from *J. mandshurica* (fig. 1*c*).

A Principal Component Analysis (PCA) of the same 19,795 high-quality SNPs as used in the global STRUCTURE analysis demonstrated striking structure within *Juglans*. The *Cardiocaryon* and *Rhysocaryon* samples segregated along PC1, whereas section *Juglans* samples were between these sections along PC1, but distinct along PC2 (fig. 1*b*). PCA plots from simulated SNP genotypes suggest that hybrid taxa are intermediary to their parents along PC1 and distinct along PC2 (Sefc and Koblmueller 2016), supporting that the *J. regia*/*J. sigillata* lineage is an ancient hybrid between *Cardiocaryon* and *Rhysocaryon*.

Phylogenetic-Network Inference

PhyloNet analyses (Yu and Nakhleh 2015), which infer species phylogenies by accounting for both incomplete lineage sorting and hybridization, sorted the samples into three major clades (fig. 2*a*), corresponding to sections *Juglans*, *Rhysocaryon*, and *Cardiocaryon*. In all cases allowing 1, 2, or 3 past hybridization events, section *Juglans* was invariably identified as a reticulate node (supplementary fig. S4, Supplementary Material online), consistent with the results of STRUCTURE and PCA. The inheritance probabilities showed that the ancestral lineage of *J. regia*/*J. sigillata* had a genomic contribution of 53% from *Cardiocaryon* and 47% from *Rhysocaryon*, slightly different from the estimates obtained with STRUCTURE (33% from *Cardiocaryon*, 67% from *Rhysocaryon*), probably because the analysis in PhyloNet was based on single-copy genes, whereas the STRUCTURE analysis was based on genome-wide SNPs.

Tests of Interspecific Gene Flow

Using the software HyDe (Blischak et al. 2018), which detects genome-scale hybridization by using phylogenetic invariants, we detected a significant signal of hybridization in both *J. regia* and *J. sigillata* through using a total of 306,457,168 bp of non-missing sites at both the population and individual level (supplementary table S4, Supplementary Material online). Furthermore, any member of section *Cardiocaryon* could have been at the base of one parental lineage and any member of section *Rhysocaryon* at the base of the other. The estimated probability (γ) of inheritance from an ancestor of section *Cardiocaryon* ranged from 0.420 to 0.449 (supplementary table S4, Supplementary Material online), whereas the proportion of ancestry from *Rhysocaryon* was somewhat higher, as also inferred in the STRUCTURE analyses (above). Both HyDe and STRUCTURE rely on genome-wide SNP data. Each individual of *J. regia* and *J. sigillata* also showed significant levels of hybridization, with the inheritance probability from *Cardiocaryon*, γ, ranging from 0.414 to 0.458 (fig. 3*a*),

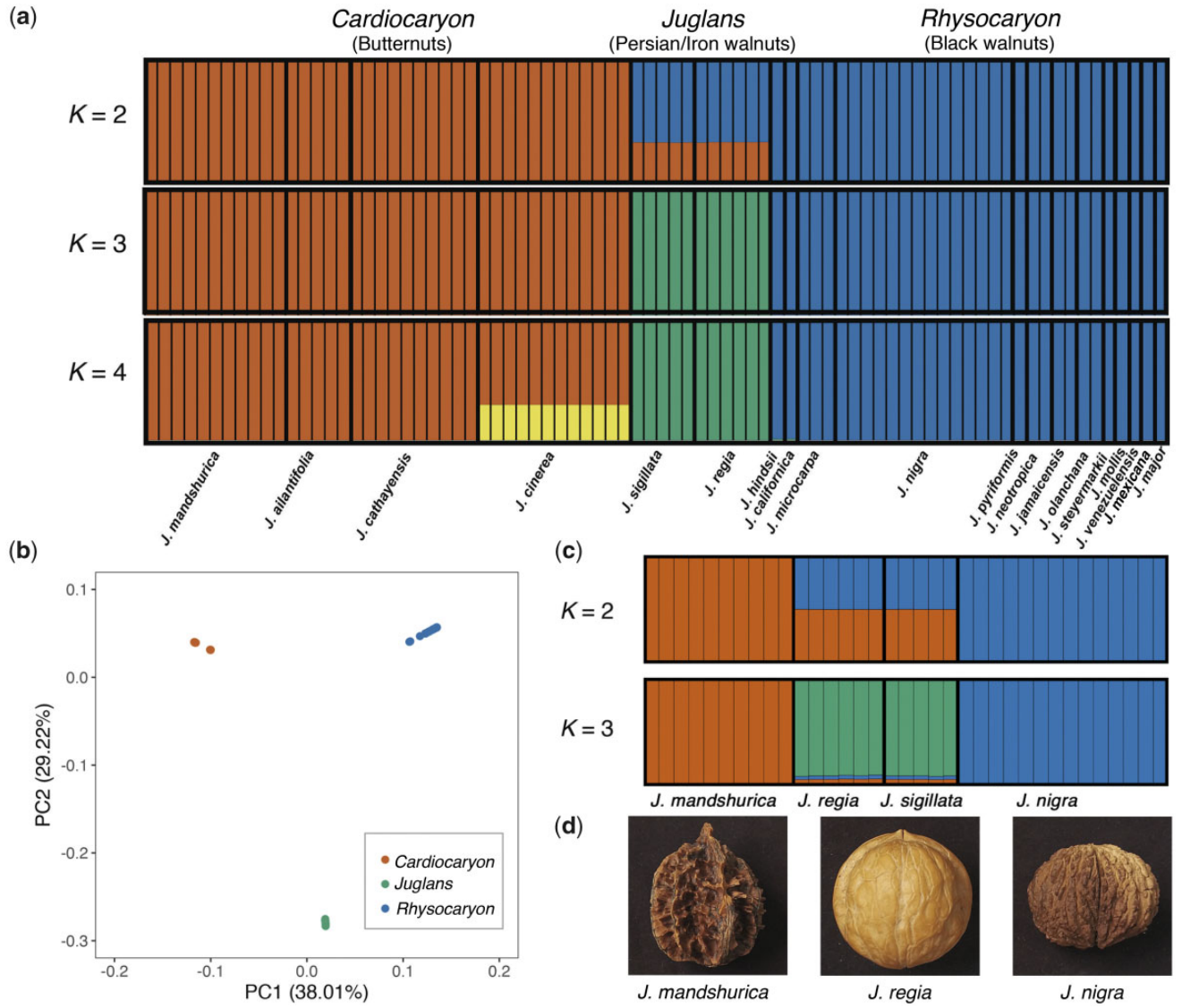

FIG. 1. Population structure based on independent nonmissing SNPs. (*a*) Results of STRUCTURE analyses of 80 individuals of *Juglans*. (*b*) PCA of the same 80 individuals. (*c*) Assignment probabilities for *Juglans regia* and *J. sigillata* individuals in STRUCTURE analyses when using population information of *J. mandshurica* and *J. nigra*. (*d*) Nut photos of *J. mandshurica*, *J. nigra*, and *J. regia*.

indicating that all sampled individuals of *J. regia* and *J. sigillata* were admixed.

Currently, the most widely used ABBA-BABA test for hybridization detection is based on counts of ancestral (A) and derived (B) alleles in sets of four taxa with known phylogenetic relationships (here, *J. regia* [or *J. sigillata*], *J. mandshurica* and *J. nigra*, and *Pl. strobilacea*). The test statistic *D* does not differ significantly from 0 when the derived alleles in *J. nigra* match alleles in *J. mandshurica* and *J. regia* equally often, or when the derived alleles in *J. mandshurica* match alleles in *J. nigra* and *J. regia* (or *J. sigillata*) equally often. The ABBA-BABA tests showed that *J. nigra* is significantly closer to *J. regia* or *J. sigillata* than to *J. mandshurica*, and similarly, *J. mandshurica* is significantly closer to *J. regia* or *J. sigillata* than to *J. nigra* (table 1), indicating that *J. regia*/*J. sigillata* likely originated as a hybrid between black walnuts and Asian butternuts.

Another ABBA-BABA test in which we included *J. cinerea*, *J. nigra*, and *J. mandshurica* revealed gene flow from black walnuts and Asian butternuts to *J. cinerea* (table 1), although HyDe failed to detect this hybridization signal (supplementary table S4, Supplementary Material online).

The Population History of *J. regia* and *J. sigillata*

The analyses described so far provide evidence for ancient hybridization having played a role in the origin of the Persian walnut lineage (*J. regia*/*J. sigillata*) and raise the question of the direction and timing of the genetic exchange. There are four possible scenarios: a hybrid origin, lineage merging, gene flow from *J. nigra*, or gene flow from *J. mandshurica* (fig. 4*a*). We used an ABC approach to determine which of the four scenarios best fit the available data for Persian walnut and *J. sigillata*. Because postdivergence gene flow has occurred between sections *Cardiocaryon* and *Rhysocaryon* (supplementary section S2 and fig. S5 and table S5, Supplementary Material online) but is difficult to model, we did not use summary statistics for *J. mandshurica* and *J. nigra* in our

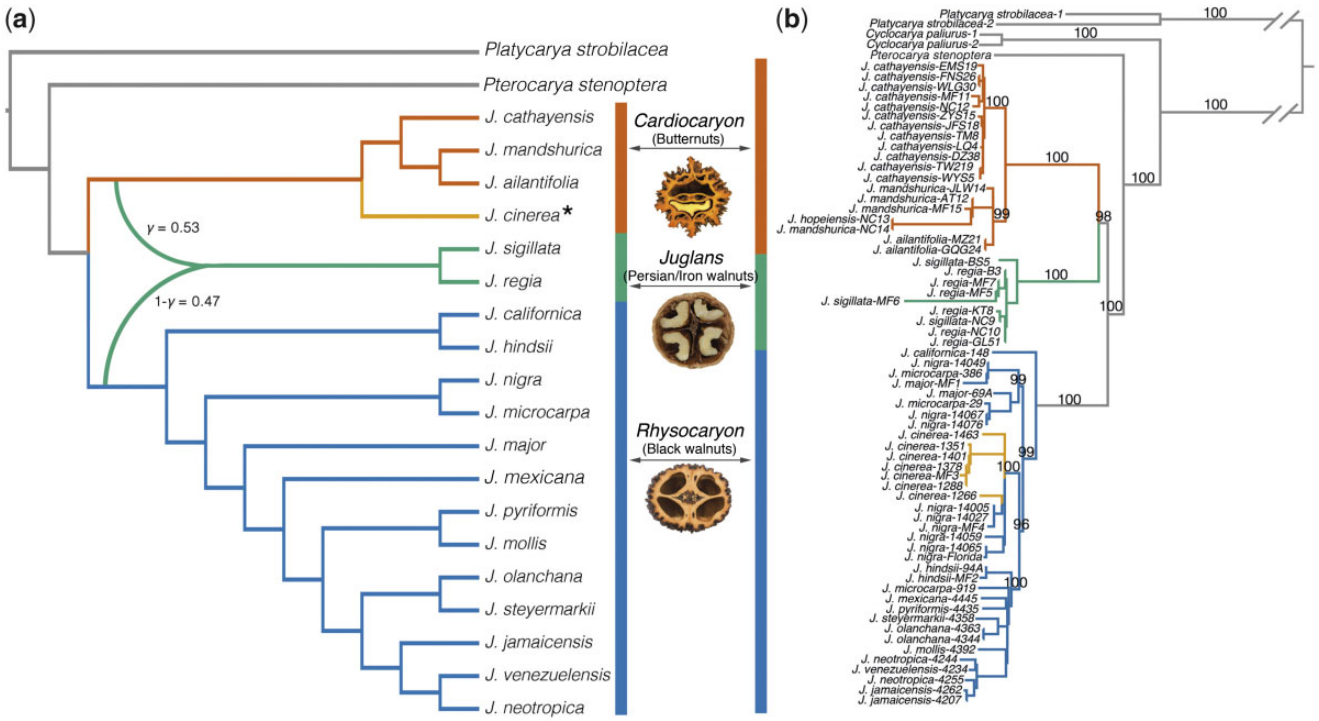

FIG. 2. Species network inference and chloroplast phylogeny. (*a*) Optimal species network inferred using the PhyloNet software. The result is maximum pseudo-likelihood tree with one reticulation. The γ value indicates the inheritance probability from the ancestor of sect. *Cardiocaryon*, whereas $1-\gamma$ indicates that from the ancestor of sect. *Rhysocaryon*. (*b*) A ML phylogeny from 68 entire plastid genomes of *Juglans* and two outgroups, with bootstrap support values $\geq$96 above branches. The asterisk marking *Juglans cinerea* refers to the only case of phylogenomic discord between the nuclear and plastid topologies (see Discussion).

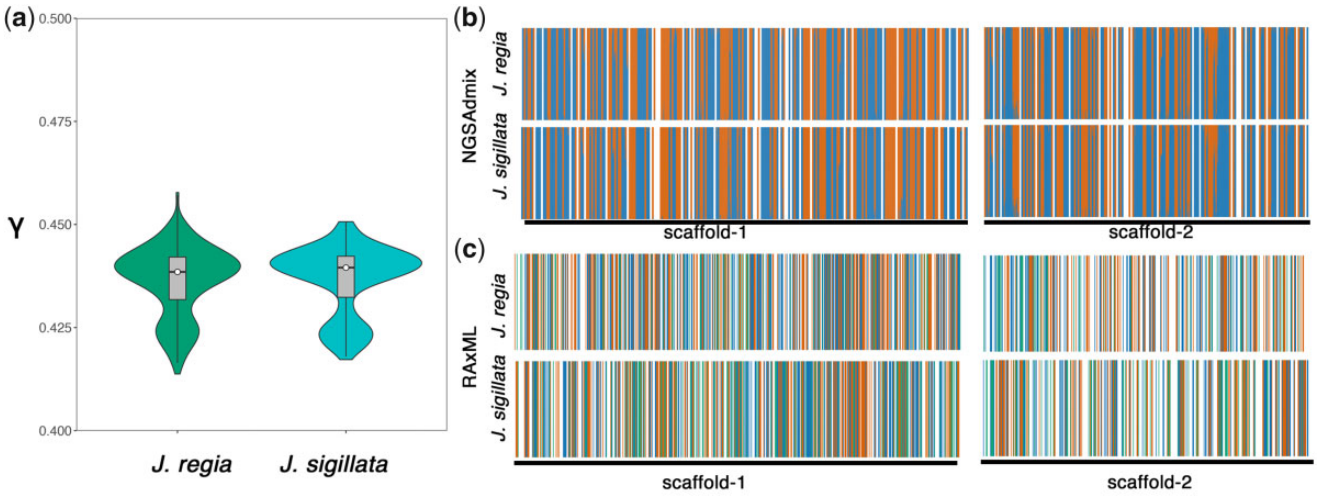

FIG. 3. HyDe individual test and stepping-windows admixture and phylogenetic analyses for *Juglans regia* and *J. sigillata*. (*a*) Violin plots of the distribution of γ with HyDe for six individuals of *J. regia* and five individuals of *J. sigillata*. γ indicates the inheritance probabilities from the parental species of sect. *Cardiocaryon* to the inferred individual. (*b*) Results of the NGSADMIX-analysis assignments for 10-kb nonoverlapping windows across scaffolds 1 and 2 of *J. regia* and *J. sigillata*. The windows show *J. regia* windows clustering with *J. mandshurica* in red, *J. regia* windows clustering with *J. nigra* in blue, and *J. regia* clustering by itself in white. (*c*) RAxML analysis on 10-kb windows depicting whether *J. regia* or *J. sigillata* grouped monophyletically with *J. mandshurica* (red), *J. nigra* (blue), formed their own clade (green), or remained unresolved (white).

modeling. Instead, we used only polymorphism information in *J. regia* and *J. sigillata* such that any posthybridization gene flow between *J. mandshurica* and *J. nigra* will not affect our tests (Materials and Methods). The gene-flow event was dated back to sometime during the Pleistocene (supplementary table S5, Supplementary Material online), and we presume it was posterior to the hybridization event leading to the origin of Persian walnut.

The best-fitting model was a lineage merging model (posterior probability = 0.783 ± 0.047, supplementary fig. S6, Supplementary Material online). The divergence between black walnuts and butternuts was estimated to have occurred

Table 1. Results from Patterson's *D* Test for Introgression between Species with 500-kb Block Jack-Knifed SE Estimates and Significance Values.

	SP1	SP2	SP3	Patterson's *D*	Jack-Knifed SE	*Z*-Score	*P*
Juglans regia	*Juglans mandshurica*	*Juglans regia*	*Juglans nigra*	0.079	0.0017	45.50	<0.0001
	Juglans nigra	*Juglans regia*	*Juglans mandshurica*	0.110	0.0017	61.00	<0.0001
Juglans sigillata	*Juglans mandshurica*	*Juglans sigillata*	*Juglans nigra*	0.106	0.0014	67.73	<0.0001
	Juglans nigra	*Juglans sigillata*	*Juglans mandshurica*	0.097	0.0017	55.90	<0.0001
Juglans cinerea	*Juglans mandshurica*	*Juglans cinerea*	*Juglans nigra*	0.029	0.0014	19.21	<0.0001
	Juglans nigra	*Juglans cinerea*	*Juglans mandshurica*	0.635	0.0017	359.20	<0.0001

NOTE.—Three taxa (SP1, SP2, and SP3) and an outgroup species (*Platycarya strobilacea*) were used to calculate Patterson's *D* in each computation.

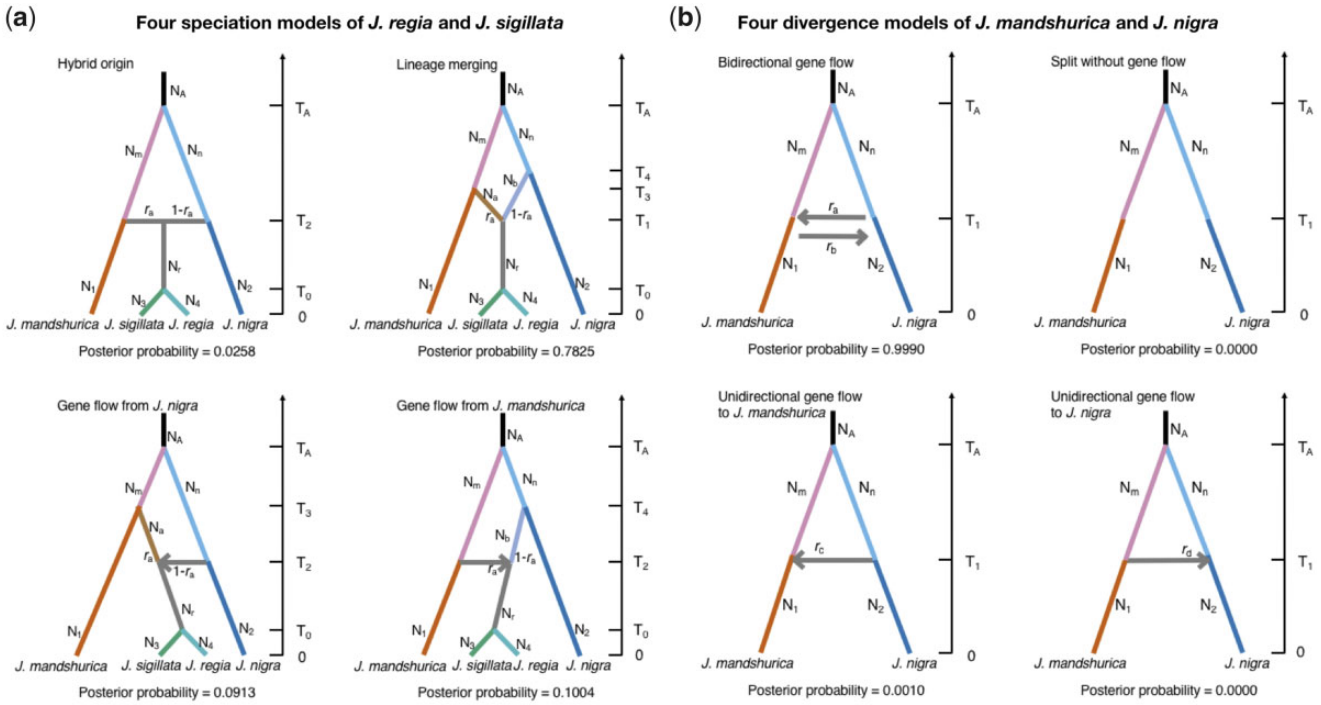

FIG. 4. ABC models and their corresponding posterior probabilities for the origin of *Juglans regia* and *J. sigillata* (*a*) and models for the divergence of *J. mandshurica* and *J. nigra* (*b*). The posterior probability of each model was estimated with logistic regression using 1% of simulated data closest to the observed data set in the model selection procedure.

~37.5 My (95% HPD: 12.9–58.8 My), comparable to an estimate obtained after excluding *J. regia*/*J. sigillata* (fig. 4*b*). The divergence time between a ghost butternut lineage and *J. mandshurica* was 20.9 My and that between a ghost black walnut lineage and *J. nigra* 23.8 My (supplementary table S6, Supplementary Material online). However, the posterior distributions for these two parameters were similar to the prior distributions, implying that the signal in the data from only *J. regia* and *J. sigillata* is insufficient for solid estimation of the divergence times. The hybridization event that gave rise to the *J. regia*/*J. sigillata* lineage was estimated to have occurred during the late Pliocene, 3.45 My (95% HPD: 1.20–8.22 My; supplementary table S6, Supplementary Material online). The estimated model parameters indicated that 47% of the nuclear genome of *J. regia*/*J. sigillata* was derived from an ancestral butternut and 53% from black walnut, similar to the above estimates obtained with HyDe and STRUCTURE.

We also did an ABC analysis of the four speciation models (supplementary fig. S7, Supplementary Material online) by allowing gene flow between *J. mandshurica* and *J. nigra* (supplementary section S3, Supplementary Material online) and obtained very similar results. Most notably, the support for the best-fitting lineage merging model was even higher (posterior probability = 0.944, supplementary fig. S7, Supplementary Material online) and the hybridization event was estimated to have occurred during the late Pliocene, 3.72 My (95% HPD: 0.87–11.93 My; supplementary table S7, Supplementary Material online), closely matching the above estimate of ~3.45 My without consideration of gene flow. To test the accuracy of parameter estimation under the lineage merging model, we used fastsimcoal2 (Excoffier et al. 2013), a continuous-time coalescent simulator. With this approach, the hybridization event was dated to at ~3.17 My, matching to ABC results (supplementary section S6, Supplementary Material online).

The Genomic Constitution of the *J. regia*/*J. sigillata* Lineage

Genome-wide population-genetic parameter estimates reveal that *J. regia*/*J. sigillata* is intermediate between *J. nigra* and

Table 2. Mean Values of Population Genomic Statistics for 10-kb Sliding Windows with 2,500-bp Steps across the Genome of *Juglans mandshurica* (*M*), *J. nigra* (*N*), *J. regia* (*R*), and *J. sigillata* (*S*).

Parameter	Species	Mean ± SD	Species	Mean ± SD
F_{st}	*M*–*N*	0.6028 ± 0.1489	*M*–*S*	0.5936 ± 0.1526
	M–*R*	0.5847 ± 0.1515	*N*–*S*	0.5883 ± 0.1520
	N–*R*	0.5834 ± 0.1505	*R*–*S*	0.1142 ± 0.0759
d_f	*M*–*N*	0.0031 ± 0.0040	*M*–*S*	0.0029 ± 0.0038
	M–*R*	0.0028 ± 0.0037	*N*–*S*	0.0030 ± 0.0040
	N–*R*	0.0029 ± 0.0039	*R*–*S*	0.0002 ± 0.0006
d_{xy}	*M*–*N*	0.0066 ± 0.0073	*M*–*S*	0.0058 ± 0.0067
	M–*R*	0.0058 ± 0.0067	*N*–*S*	0.0064 ± 0.0072
	N–*R*	0.0064 ± 0.0072	*R*–*S*	0.0025 ± 0.0034
π	*M*	0.0023 ± 0.0030		
	N	0.0028 ± 0.0034		
	R	0.0022 ± 0.0031		
	S	0.0021 ± 0.0030		

NOTE.—d_f, density of fixed differences.

J. mandshurica, although this is nonsignificant (table 2), possibly due to extensive gene flow between sections *Cardiocaryon* and *Rhysocaryon* (supplementary section S2 and fig. S5 and table S5, Supplementary Material online). The global average differentiation (F_{st}, d_{xy}) of *J. nigra* from *J. mandshurica* was higher than that between *J. regia*/*J. sigillata* and either *J. nigra* or *J. mandshurica*.

Admixture analysis NGSAdmix (Skotte et al. 2013) and maximum likelihood (ML) trees from 10-kb nonoverlapping windows among all the *J. regia* or *J. sigillata* individuals demonstrated a discordant evolutionary history throughout the hybrid genome of *J. regia* and *J. sigillata*. Thus, admixture analysis revealed 5,155 windows in which *J. regia* grouped with *J. mandshurica* and 6,361 windows in which it grouped with *J. nigra*, but only 718 windows assigned uniquely to *J. regia*, with very similar results for *J. sigillata* (fig. 3*b*). In the ML trees (with an 80% bootstrap threshold), there were 3,919 windows where *J. regia* individuals grouped with *J. mandshurica* and 4,253 windows where they grouped with *J. nigra*, but only 663 windows where *J. mandshurica* and *J. nigra* grouped together (fig. 3*c*). ML trees from 25- or 50-kb window sizes and 50%, 80%, or 90% bootstrap thresholds, and all corresponding analyses using *J. sigillata* instead of *J. regia*, gave broadly similar results (supplementary table S8, Supplementary Material online).

Phylogenomic Discord in the American Butternut, *J. cinerea*

Three clades corresponding to sections *Juglans*, *Cardiocaryon*, and *Rhysocaryon* were evident in a ML tree obtained from whole-chloroplast genome data, except that *J. cinerea* was nested near *J. nigra* within the black walnut clade (section *Rhysocaryon*), instead of the butternut clade (section *Cardiocaryon*; supplementary section S4, Supplementary Material online, fig. 2*b*). Cytonuclear discordance in *J. cinerea* has been noted before (Stanford et al. 2000; Stone et al. 2009; Dong et al. 2017) and chloroplast capture has been proposed as a plausible explanation (Aradhya et al. 2007). The ABBA-BABA test revealed gene flow between *J. cinerea* and *J. mandshurica* or *J. nigra*, suggesting that *J. cinerea* likely originated from hybridization. Among well-resolved 10-kb nonoverlapping windows (with >80% bootstrap support), 12,575 have a topology of ([*J. cinerea*, *J. mandshurica*], *J. nigra*) and only 198 of ([*J. cinerea*, *J. nigra*], *J. mandshurica*), which implies that most of the *J. cinerea* nuclear genome came from Asian butternuts. Such strong asymmetrical hybridization possibly also explains the failure of HyDe to detect hybridization in *J. cinerea* (table 1 and supplementary table S4, Supplementary Material online), because its statistical power decreases with increasing asymmetry in parental contributions to the ancestry of a hybrid lineage (Kubatko and Chifman 2015).

Once again, we relied on an ABC approach to determine the evolutionary history of *J. cinerea*. In this case, gene flow between Asian *J. mandshurica* and American *J. nigra* cannot be ignored because it must be intermingled with the hybridization event resulting in the formation of *J. cinerea*. Depending on the timing of the gene-flow event relative to the hybridization event, a total of 12 scenarios (supplementary section S5 and fig. S8, Supplementary Material online) can be constructed on top of the four basic speciation models as represented in figure 4. Two scenarios (Model 6 and Model 10), very similar in nature, were the most likely models of origin (supplementary section S5 and fig. S8, Supplementary Material online). In Model 10 (posterior probability = 0.631 ± 0.076), introgression from *J. mandshurica* contributed about 0.73 of the *J. cinerea* genome and the time for the origin of *J. cinerea* was estimated as 0.57 My (95% HPD: 0.18–1.03 My; supplementary table S9, Supplementary Material online); In Model 6 (posterior probability 0.301 ± 0.087), introgression from a distinct ghost lineage of *J. mandshurica* contributed about 0.92 of the ancestry of the *J. cinerea* genome and the origin of *J. cinerea* was dated to 0.89 My (95% HPD: 0.18–2.00 My; supplementary table S9, Supplementary Material online). Both models suggest massive introgression of nuclear DNA from an Asian butternut parent (Model 10: *J. mandshurica*; Model 6: a distinct ghost) to a ghost black walnut lineage diverged from *J. nigra* at about 7.3∼11.7 My. The only significant difference between the two scenarios concerns the butternut parent. The best-fitting model with fastsimcoal2 was Model 6 (supplementary section S6 and table S15, Supplementary Material online). Note that both models date the origin of *J. cinerea* back to the Early-Middle Pleistocene transition around 0.6–0.9 My (supplementary table S9, Supplementary Material online).

Discussion

In the present study, all genome-wide analyses converged to provide unambiguous evidence of hybridization at the roots of both the American butternut (*J. cinerea*) and the Persian walnut (*J. regia*/*J. sigillata*). The inferred late Pliocene hybrid event that gave rise to the Persian walnut around ∼3.45 My fits with the absence of this species in the ancient European fossil record (van der Ham 2015). The genomic mosaicism of both the Persian walnut (incl. *J. sigillata*) and the American butternut *J. cinerea* provides evidence for the importance of hybridization throughout the evolution of walnuts, matching

recent results from subsets of Chinese walnut species (Bai et al. 2016; Yuan et al. 2018; Zhao et al. 2018). However, our results from the entire genus, which today is most species-rich in North and South America while Europe has a single species (supplementary fig. S1, Supplementary Material online), may be the first example of inferred hybridization and introgression among extinct species at both shallow and deep evolutionary timescales.

The placement of the American butternut, *J. cinerea*, with the Asian butternuts in nuclear data but within the American section *Rhysocaryon* in a plastid phylogeny (fig. 2*b*; also seen with more limited gene and species sampling in the studies of Stanford et al. (2000), Aradhya et al. (2007), and Dong et al. (2017)), can also now be understood as the results of hybridization. Cytonuclear discordance of this kind is common in plants and is usually explained by chloroplast capture (Rieseberg and Soltis 1991; Wolfe and Elisens 1995; Folk et al. 2017). Our ABC analysis instead inferred that the evolution of *J. cinerea* is due to massive introgression from an immigrating Asian butternut into the genome of an American black walnut (supplementary section S5 and fig. S8, Supplementary Material online). This parsimoniously explains why *J. cinerea* has the nuclear genome of an Asian butternut and the chloroplast genome of an American black walnut. Such pollen swamping, best documented in oaks (Petit et al. 2004), has also been inferred in *Mercurialis annua* (Christelová et al. 2011). One might therefore return to classifying the "hybrid species" *J. cinerea* in its own section, *Trachycaryon* (Manning 1978), thereby highlighting its complex evolutionary history.

Based on walnut fruit fossils from Western North America dating to 48.32 My, *Pterocarya* and *Juglans* (which have very different fruits) had diverged from each other by the late early Eocene, followed by an initial split into black walnuts and butternuts, probably during the middle Eocene (~45 My) in North America (Manchester 1987, 1994). By the late Oligocene, walnuts must have expanded from America to Europe, probably both via the Beringian land bridge and the North Atlantic land bridge. Evidence for a North Atlantic crossing comes from *Juglans* pollen records from Axel Heiberg Island (middle Eocene, ca. 45 My) and Svalbard (late Eocene) (McIntyre 1991). Spread of a member of *Cardiocaryon* (perhaps via the Beringian land bridge) to Asia and further expansion to Europe would explain the fossil remains of butternut-type fruits (called *J. bergomensis*) in the Miocene of western Washington State and the Pliocene of Europe (Manchester 1987; Martinetto 2015; van der Ham 2015; Smith and Manchester 2018), as well as the presence of *Pterocarya* or *Cardiocaryon* pollen in the Eocene of Messel near Frankfurt (Manchester 1994).

The cooling climate of the Upper Pliocene may have led to range shifts of the butternut and black walnut lineages in Eurasia, permitting the contact required for the hybridization that gave rise to Persian walnut, *J. regia* of which *J. sigillata* is an ecotype maintained as a landrace (Zhao et al. 2018). In America and Asia, black walnut and butternut lineages survived and formed today's species; in Europe, by contrast, walnut diversity was lost during the Pleistocene climate oscillations, with eventually only the newly formed hybrid lineage surviving, probably in southern refugia. Its hybrid origin may also have resulted in adaptive introgression, which may have helped Persian walnuts to survive. Evidence for introgression is being documented in an increasing number of systems, though demonstrating the adaptive function of introgressed genomic regions remains difficult (Jones et al. 2018; Taylor and Larson 2019).

Homoploid hybrid origin of a new species is not equivalent to hybrid speciation (Schumer et al. 2014) because the latter requires hybridization per se to cause reproductive isolation of the hybrid lineage from both parents for which we have no evidence in walnuts. Our ABC analysis supports that the Persian walnut has a hybrid origin, but we know little about the specific isolating mechanisms derived from hybridization itself. Today, Persian walnut can readily hybridize with both black walnut and butternuts (Xu et al. 2007; Woeste and Michler 2011; Shu et al. 2016), suggesting no inherent reproductive barriers with congeners. Although the number of cases of homoploid hybrid speciation is increasing (Sun et al. 2014; Elgvin et al. 2017; Lamichhaney et al. 2018; Ru et al. 2018), information is still lacking that would link the presence of hybrid ancestry in the genome to the process of speciation by hybridization (Schumer et al. 2014). Perhaps the criteria for homoploid hybrid speciation are too stringent, causing researchers to overlook important contributions of hybridization to evolution and speciation (Nieto Feliner et al. 2017). Whether homoploid hybrid speciation is a common speciation mechanism in general therefore remains an outstanding question (Comeault and Matute 2018).

Regardless of the specific biogeographic scenario, our genome-wide data from 80 individuals representing 19 walnut species and two outgroup genera provide clear evidence that the Persian walnut arose as a hybrid between American and Asian lineages, and the results further resolve the controversy concerning the American butternut, *J. cinerea*, which turns out to result from massive nuclear gene introgression involving an American black walnut through pollen swamping by an immigrating Asian butternut. Gene flow, in the form of hybridization and introgression, is increasingly recognized as a contributor to the evolution of organisms. We now have good evidence from the ancient history of walnuts and butternuts, and similar events may have influenced the evolution of other wind-pollinated temperate trees.

Materials and Methods

Sampling Design

We sampled 80 individuals from 19 species of *Juglans*, namely six North American temperate species, *J. californica*, *J. cinerea*, *J. hindsii*, *J. major*, *J. microcarpa*, and *J. nigra*; six Central American subtropical species, *J. jamaicensis*, *J. mexicana*, *J. mollis*, *J. olanchana*, *J. pyriformis*, and *J. steyermarkii*; two South American tropical species, *J. neotropica* and *J. venezuelensis*; three Asian species, *J. ailantifolia*, *J. cathayensis*, and *J. mandshurica* and the Eurasian entities *J. regia* and *J. sigillata*. As outgroup species we sampled *P. stenoptera* and *Pl. strobilacea* (supplementary table S2, Supplementary

Material online, provides information on taxonomic authors and herbarium vouchers). The Asian butternut *J. hopeiensis* was only included in phylogenetic analyses of plastid data. We did not sample *J. australis* Griesb. from Argentina/Bolivia and *J. boliviana* (C.DC.) Dode from Bolivia; *J. soratensis* W.E. Manning, also from Bolivia, is considered a hybrid and was also not included.

Reference Genome Assembly

Fragment libraries of 250, 350, and 450 bp were sequenced with a paired-end 150-bp strategy on the Illumina HiSeq × ten sequencing platform at a depth of 60×, 45×, and 58× for *P. stenoptera*. A fragment library of 350 bp was similarly sequenced at a depth of 53× for *Pl. strobilacea*. Jumping libraries with insert sizes of 3k, 5k, and 10k bp for *P. stenoptera* and 2k, 5k, and 10k for *Pl. strobilacea* were sequenced with a paired-end 150-bp strategy on Illumina HiSeq X ten. The total depth of jump libraries was about 100× and 64× for *P. stenoptera* and *Pl. strobilacea*, respectively. All libraries were assembled and scaffolded using allPathsLG version 474117 (Gnerre et al. 2011) with default parameters after filtering reads containing sequence adaptors or reads not paired properly. We also predicted and annotated genes and other features of each genome (supplementary section S1, Supplementary Material online).

Sequencing, Reads Mapping, and Variant Calling

Whole-genome sequencing using paired-end libraries with an insert size of 350 bp was performed on Illumina Hiseq X-ten instruments with 150-bp read length on each end by NovoGene (Beijing, China). Samples were sequenced to an average depth of 30×. Because *P. stenoptera* is equally related to all Juglans species, the reads from 80 individuals of Juglans were mapped to the *P. stenoptera* reference genome. SNPs from each individual were called and joined to create a multisample SNP data set using SENTIEON DNAseq software packages v. 201711.05 (Weber et al. 2016). To control the quality of SNPs, triallelic and tetra-allelic SNP sites or sites with missing data or a mapping depth <10× or >60× in any individuals were removed. Then we use a Q20 filter and called a heterozygous genotype if the depth of an allele was 20× to 60× and the proportion of a nonreference allele was between 20% and 80%, or if the depth was 10× to 20× and the proportion of a nonreference allele was between 10% and 90%, otherwise a homozygous genotype was called (Nielsen et al. 2011). After filtering, a total number of 3,998,064 SNPs remained. These were further thinned using a distance filter of interval >5k-bp and a rare SNP filter of MAF > 0.05. The final data set contained 19,795 SNPs (supplementary section S1, Supplementary Material online).

Population Structure

We used STRUCTURE v. 2.3.4 (Pritchard et al. 2000) to cluster individuals based on $K = 1$–5 using the admixture model and uncorrelated allele frequencies. To account for unequal sample sizes among species, we set POPALPHAS to 1, with an initial value of alpha = 0.25 as suggested by Meirmans (2019). Then, we choose *J. mandshurica* and *J. nigra* as the representatives for sections *Cardiocaryon* and *Rhysocaryon*, respectively and did another assignment for the putative hybrid species *J. regia* and *J. sigillata*. After the same filtering strategy, we used a total of 26,995 SNPs to conduct an assignment test on $K = 2$–3 for *J. regia* and *J. sigillata* with USEPOPINFO = 1 for *J. mandshurica* and *J. nigra*. Both assignments were performed ten times to ensure a stable result, and Markov Chain Monte Carlo analyses were run for 50,000 iterations, after a burnin period of 20,000 iterations. A PCA was run on the 19,795 SNP data set using the *R* package SNPRelate v. 1.6.2 (Zheng et al. 2012) with default settings.

Phylogenetic-Network Analysis

To obtain single-copy genes, we mapped reads of each individual to a closely related reference genome using BWA v. 0.7.12 (Li and Durbin 2009). Species of section *Rhysocaryon* were mapped onto the *J. nigra* genome, species of section *Cardiocaryon* onto the *J. mandshurica* genome, and those of section *Juglans* onto the *J. regia* genome (genome version V3, available in http://cmb.bnu.edu.cn/juglans/ or https://www.ncbi.nlm.nih.gov/bioproject/PRJNA356989/). Single-copy genes ($N = 2{,}901$) were chosen to perform the nuclear phylogenetic analysis (supplementary section S1, Supplementary Material online). RAxML v. 8.2.8 (Stamatakis 2014) was used to build a ML gene tree for each gene under the GTR + GAMMA substitution model, using the rapid-bootstrapping approach implemented in RAxML (100 bootstrap replicates), and with *Pl. strobilacea* set as outgroup. Species networks that modeled incomplete lineage sorting and gene flow using a pseudo-maximum likelihood approach (Yu and Nakhleh 2015) were carried out with PHYLONET v. 3.6.1 (Than et al. 2008) with the command "InferNetwork_MPL" and using the individual gene trees. Network searches were performed using only nodes in the rooted ML gene trees that had a bootstrap support of at least 75%, allowing for 0–3 reticulations, and optimizing the branch lengths and inheritance probabilities of the returned species networks under pseudo-likelihood.

Tests of Interspecific Gene Flow

We used HyDe (Blischak et al. 2018), a software package for detecting hybridization from phylogenetic invariants that arise under the coalescent model with hybridization, to detect hybridization in the consensus genomes built using SNPs. Similar to ABBA-BABA tests, HyDe considers a rooted, four-taxon network consisting of an outgroup and a triplet of ingroup populations. The distribution of site patterns was used to infer a hybrid ingroup population that with probability γ is sister to one population ($P1$) and with probability $1 - \gamma$ sister to the other population ($P2$). The null hypothesis was that when admixture was absent, the expected value of γ should be 0. We perform HyDe tests for 19 *Juglans* species (not including *J. hopeiensis*) with *Pl. strobilacea* as an outgroup. We assumed each species to be a hybrid species, resulting in $\binom{19}{3} * 3 = 2,907$ hypothesis tests at the species level. At the individual level, significant hybridization was evaluated for *J. regia* and *J. sigillata* (a total of 11 individuals) where

the two parental species could be any of four species of sect. *Cardiocaryon* and any of 13 species of section *Rhysocaryon*, resulting in $4 \times 11 \times 13 = 572$ hypothesis tests. Results from those runs were filtered with 1% critical value on *Z*-scores.

We used the *D*-statistic (also known as the ABBA-BABA test) (Green et al. 2010; Durand et al. 2011) to test for the possibility of admixture between *J. nigra* and *J. mandshurica*, *J. nigra* and *J. regia*, and *J. mandshurica* and *J. regia*. The analysis uses patterns of ancestral and derived alleles in the ingroups and outgroups to distinguish between incomplete lineage sorting and hybridization. The *D*-statistic has been shown to be a robust, although conservative, method for identifying introgressed loci. Whole-genome *D*-statistics were calculated for two topologies ([*J. nigra*, *J. regia*], *J. mandshurica*) and ([*J. mandshurica*, *J. regia*], *J. nigra*) using ANGSD's ABBA BABA multipopulation tool (Soraggi et al. 2018) with *Pl. strobilacea* as the outgroup. We chose five individuals from each *Juglans* species to test for significant evidence of admixture using a weighted block jackknife with 500-kb nonoverlapping blocks. We considered *Z*-scores >3 to be significant. We also did the same ABBA-BABA test in *J. sigillata*, to test the possibility of admixture between *J. nigra* and *J. mandshurica*, *J. nigra* and *J. sigillata*, and *J. mandshurica* and *J. sigillata*.

Tests of Population History by ABC Modeling

We compared four hypothesized models of speciation (fig. 4*a* and supplementary table S10, Supplementary Material online) through analysis of the nuclear data set using DIYABC v. 2.1 (Cornuet et al. 2014) for *J. nigra*, *J. mandshurica*, *J. regia*, and *J. sigillata*. In all four models, T_A was the divergence time between ancestral populations of *J. nigra* and *J. mandshurica* and NA was the effective population size of the common ancestor of those species; T_0 was the divergence time between two ancestral populations of *J. regia* and *J. sigillata*, and *Nr* was the effective population size of the ancestral population of section *Juglans* (*J. regia* and *J. sigillata*). Model 1 (hybrid origin) assumed an admixture event between *J. mandshurica* (*ra*) and *J. nigra* ($1 - ra$) giving birth to section *Juglans* at T_1. Model 2 (lineage merging) assumed two speciation events that resulted in lineages sister to *J. mandshurica* and *J. nigra* at T_2 and T_3, respectively, which then came together to form section *Juglans* at T_1. The last two demographic models have the same parameters and differ only in their topology: in Model 3 (gene flow from *J. nigra*), *J. mandshurica* and a progenitor lineage diverged from the common ancestor of butternuts at T_4, and then *J. nigra* was introgressed into this progenitor lineage at T_1 and formed section *Juglans*; in Model 4 (gene flow from *J. mandshurica*), *J. nigra* and a progenitor lineage diverged from the common ancestor of black walnuts at T_4, and then *J. mandshurica* was introgressed into this progenitor lineage at T_1 and formed section *Juglans*. We performed four million simulations for the 26,995 SNP data set with a 5-kb interval and MAF > 0.05 criterion from 35 individuals from the four species. In order to avoid the influence of gene flow between *J. mandshurica* and *J. nigra*, we only used four one-sample summary statistics from *J. regia* and *J. sigillata* (genic diversities: proportion of zero values, mean of nonzero values, variance of nonzero values, and mean of complete distribution), and four two-sample summary statistics between *J. regia* and *J. sigillata* only (F_{st}: mean and variance of nonzero values; Nei's distances: proportion of zero values; mean of complete distribution). The 40,000 (1%) simulated data sets closest to the observed data set were selected for choosing the best scenario by logistic regression. Then, we evaluated five million data sets to determine the best model and estimated posterior parameter distributions by using 50,000 (1%) simulated data sets closest to the observed data set.

Population Genomic Analyses in Sliding Windows

Estimates for F_{st} were calculated in overlapping sliding windows (10 kb in size with 2.5-kb steps) with ANGSD (Korneliussen et al. 2013) with -gl1. We calculated F_{st} as the weighted mean for each window across the genome. To calculate the nucleotide diversity within populations (π), density of differences (d_{xy}), and fixed differences (d_f), we used consensus genomes built using SNPs. Numbers of differences were divided by the number of nonmissing sites in every 10-kb sliding window with 2.5-kb steps to obtain the diversity value per site. In order to infer the linkage disequilibrium in three focal species, we used Beagle v. 4.1 (Browning and Browning 2007) to resolve (phase) the distinct haplotypes within each sample. Phased SNPs were further applied into VCFtools v. 0.1.13 (Danecek et al. 2011) to calculate the mean r^2 (correlation coefficient) between pairs of SNPs within 50-kb windows.

We used NgsAdmix v. 32 (Skotte et al. 2013) to estimate admixture in 10-kb stepping windows across the genomes (scaffolds with length >1 Mb) of *J. mandshurica*, *J. nigra*, and *J. regia* individuals. We first calculated genotype likelihoods from the bam files mapped to *Pl. strobilacea* in ANGSD with parameters "-doGlf2 -doMajorMinor 1 -SNP_pval 1e-6 -doMaf 1." The resulting file was then split into 10-kb stepping windows, put into NgsAdmix, and run with $K = 2$ as the number of ancestral populations. To visualize the proportions of the parent taxa's ancestry in *J. regia* for each window, we set *J. mandshurica* and *J. nigra* as parent 1 and parent 2 except for windows where two parent populations were not clearly resolved.

In addition, RAxML was run for 10-kb stepping windows across the genome for consensus sequences from *J. mandshurica*, *J. nigra*, and *J. regia* populations. Six individuals were chosen from each population to eliminate bias caused by unequal sample sizes. The substitution model was again GTR + GAMMA, with 100 rapid bootstraps, and *Pl. strobilacea* as the outgroup. The resulting trees for each window were then categorized on the basis of whether all *J. regia* individuals grouped with *J. mandshurica* or *J. nigra*, or all individuals of *J. mandshurica* and *J. nigra* grouped together or were unresolved (*J. regia* did not form a monophyletic group). The same method was used with 50- and 25-kb stepping windows to test whether window size affected the proportion of resolved phylogenies.

Chloroplast Phylogenetic Analysis

Reads from each individual were mapped against the chloroplast genome of Persian walnut NC_028617.1

(Peng et al. 2017) using the BWA-MEM algorithm (Li and Durbin 2009) of BWA v. 0.7.12. We then performed variant calling using SAMTOOLS v. 1.3 (Li 2011), with SNPs converted to the Variant Call Format. Using the annotation of the Persian walnut NC_028617 chloroplast genome, 70 protein-coding genes were aligned with MAFFT v. 7.017 (Katoh and Standley 2013) and then converted to CDS alignment with PAL2NAL v. 14 (Suyama et al. 2006). We treated the first, second, and third codon positions from each gene as different subsets, creating in total $3 \times 70 = 210$ subsets. PartitionFinder v. 2.1 (Lanfear et al. 2017) was used to partition the data into subsets of genes evolving at a similar rate and under the same nucleotide substitution model. The best partitioning scheme comprised seven subsets with lengths of 1,294–13,108 bp. Phylogenetic trees were then inferred with RAxML from the seven partitioned sequence alignments (supplementary section S4, Supplementary Material online).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We are grateful to Shou-Hsien Li and Song Ge for their insightful comments and to Pei-Chun Liao, Jie Liu, and Dun-Yan Tan for assistance with sampling. This work was supported by the National Key R&D Program of China (2017YFA0605100), the National Natural Science Foundation of China (41671040 and 31421063), the "111" Program of Introducing Talents of Discipline to Universities (B13008), and a key project of State Key Laboratory of Earth Surface Processes and Resource Ecology. The mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the US Department of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable. The authors declare no competing financial interests.

Author Contributions

W.-N.B. and D.-Y.Z. conceived of the study. W.-N.B., D.-Y.Z., and S.S.R. wrote the manuscript. B.-W.Z. and P.-C.Y. assembled the genomes. B.-W.Z., W.-N.B., L.-L.X, N.L., X.-H.J, and P.-C.Y. performed the analyses. K.L. and K.E.W. provided samples, contributed ideas, and assisted in editing the manuscript.

References

Aradhya MK, Potter D, Gao F, Simon CJ. 2007. Molecular phylogeny of *Juglans* (Juglandaceae): a biogeographic perspective. *Tree Genet Genomes* 3:363–378.

Bai W-N, Wang W-T, Zhang D-Y. 2016. Phylogeographic breaks within Asian butternuts indicate the existence of a phytogeographic divide in East Asia. *New Phytol*. 209(4):1757–1772.

Blischak PD, Chifman J, Wolfe AD, Kubatko LS. 2018. HyDe: a Python package for genome-scale hybridization detection. *Syst Biol*. 67(5):821–829.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 81(5):1084–1097.

Christelová P, Valárik M, Hřibová E, De Langhe E, Doležel J. 2011. A multi gene sequence-based phylogeny of the Musaceae (banana) family. *BMC Evol Biol*. 11:103.

Comeault AA, Matute DR. 2018. Genetic divergence and the number of hybridizing species affect the path to homoploid hybrid speciation. *Proc Natl Acad Sci U S A*. 115(39):9761.

Cornille A, Gladieux P, Smulders MJM, Roldan-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L. 2012. New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. *PLoS Genet*. 8:e100273.

Cornuet J-M, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin J-M, Estoup A. 2014. DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics* 30(8):1187–1189.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

Dong WP, Xu C, Li WQ, Xie XM, Lu YZ, Liu YL, Jin XB, Suo ZL. 2017. Phylogenetic resolution in *Juglans* based on complete chloroplast genomes and nuclear DNA sequences. *Front Plant Sci*. 8:1148.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. *Mol Biol Evol*. 28(8):2239–2252.

El Baidouri M, Murat F, Veyssiere M, Molinier M, Flores R, Burlot L, Alaux M, Quesneville H, Pont C, Salse J. 2017. Reconciling the evolutionary origin of bread wheat (*Triticum aestivum*). *New Phytol*. 213(3):1477–1486.

Elgvin TO, Trier CN, Torresen OK, Hagen IJ, Lien S, Nederbragt AJ, Ravinet M, Jensen H, Saetre GP. 2017. The genomic mosaicism of hybrid speciation. *Sci Adv*. 3(6):e1602996.

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genet*. 9(10):e1003905.

Fjellstrom RG, Parfitt DE. 1995. Phylogenetic analysis and evolution of the genus *Juglans* (Juglandaceae) as determined from nuclear genome RFLPs. *Plant Syst Evol*. 197(1-4):19–32.

Folk RA, Mandel JR, Freudenstein JV. 2017. Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Syst Biol*. 66:320–337.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 108(4):1513–1518.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz M-Y, et al. 2010. A draft sequence of the Neandertal genome. *Science* 328(5979):710–722.

Jones MR, Mills LS, Alves PC, Callahan CM, Alves JM, Lafferty DJ, Jiggins FM, Jensen JD, Melo-Ferreira J, Good JM. 2018. Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* 360(6395):1355–1358.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.

Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's *D* and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* 14:289.

Kubatko LS, Chifman J. 2015. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BioRxiv: 034348. doi: https://doi.org/10.1101/034348

Lamichhaney S, Han F, Webster MT, Andersson L, Grant BR, Grant PR. 2018. Rapid hybrid speciation in Darwin's finches. *Science* 359(6372):224–227.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models

of evolution for molecular and morphological phylogenetic analyses. *Mol Biol Evol*. 34(3):772–773.
Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27(21):2987–2993.
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.
Lu AM, Stone DE, Grauke LJ. 1999. Juglandaceae. In: Wu ZY, Raven PH, editors. Flora of China, 4, Cycadaceae through Fagaceae. Beijing (China)/St. Louis (MO): Science Press/Missouri Botanical Garden Press. p. 277–285.
Manchester SR. 1987. The fossil history of Juglandaceae. *MO Bot Garden Monogr*. 21:1–137.
Manchester SR. 1994. Fruits and seeds of the Middle Eocene Nut Beds Flora, Clarno Formation, Oregon. *Palaeontogr Am*. 58:1–114.
Manning WE. 1978. The classification within the Juglandaceae. *Ann MO Bot Garden* 65(4):1058–1087.
Manos PS, Stone DE. 2001. Evolution, phylogeny, and systematics of the Juglandaceae. *Ann MO Bot Garden* 88(2):231–269.
Martinetto E. 2015. Monographing the Pliocene and early Pleistocene carpofloras of Italy: methodological challenges and current progress. *Palaeontogr Abteilung B* 293(1-6):57–99.
McIntyre DJ. 1991. Pollen and spore flora of an Eocene forest, eastern Axel Heiberg Island. *Bull Geol Surv Can*. 403:83–97.
Meirmans PG. 2019. Subsampling reveals that unbalanced sampling affects STRUCTURE results in a multi-species dataset. *Heredity* 122(3): 276–287.
Munoz-Rodriguez P, Carruthers T, Wood JRI, Williams BRM, Weitemier K, Kronmiller B, Ellis D, Anglin NL, Longway L, Harris SA, et al. 2018. Reconciling conflicting phylogenies in the origin of sweet potato and dispersal to polynesia. *Curr Biol*. 28(8):1246.
Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet*. 12(6):443–451.
Nieto Feliner G, Alvarez I, Fuertes-Aguilar J, Heuertz M, Marques I, Moharrek F, Pineiro R, Riina R, Rossello JA, Soltis PS, et al. 2017. Is homoploid hybrid speciation that rare? An empiricist's view. *Heredity* 118(6):513–516.
Peng S, Yang G, Liu C, Yu Z, Zhai M. 2017. The complete chloroplast genome of the *Juglans regia* (Juglandales: Julandaceae). *Mitochondrial DNA A* 28(3):407–408.
Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytol*. 161(1):151–164.
Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155(2):945–959.
Rieseberg LH, Soltis DE. 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evol Trends Plants* 5:65–84.
Ru DF, Sun YS, Wang DL, Chen Y, Wang TJ, Hu QJ, Abbott RJ, Liu JQ. 2018. Population genomic analysis reveals that homoploid hybrid speciation can be a lengthy process. *Mol Ecol*. 27(23):4875–4887.
Schumer M, Rosenthal GG, Andolfatto P. 2014. How common is homoploid hybrid speciation? *Evolution* 68(6):1553–1560.
Sefc KM, Koblmueller S. 2016. Ancient hybrid origin of the eastern wolf not yet off the table: a comment on Rutledge et al. (2015). *Biol Lett*. 12(2):20150834.
Shu Z, Zhang X, Yu D, Xue S, Wang H. 2016. Natural hybridization between Persian walnut and Chinese walnut revealed by simple sequence repeat markers. *J Am Soc Horticult Sci*. 141(2):146–150.
Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195(3):693–702.
Smith M, Manchester SR. 2018. Nut of *Juglans bergomensis* (Balsamo Crivelli) Massalongo in the Miocene of North America. *Acta Palaeobot*. 58(2):199–208.
Soraggi S, Wiuf C, Albrechtsen A. 2018. Powerful inference with the *D*-statistic on low-coverage whole-genome data. *G3 (Bethesda)* 8(2):551–566.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
Stanford AM, Harden R, Parks CR. 2000. Phylogeny and biogeography of *Juglans* (Juglandaceae) based on matK and ITS sequence data. *Am J Bot*. 87(6):872–882.
Stone DE, Oh S-H, Tripp EA, Rios GL, Manos PS. 2009. Natural history, distribution, phylogenetic relationships, and conservation of Central American black walnuts (*Juglans* sect. *Rhysocaryon*). *J Torrey Bot Soc*. 136(1):1–25.
Sun Y, Abbott RJ, Li L, Li L, Zou J, Liu J. 2014. Evolutionary history of Purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Mol Ecol*. 23(2):343–359.
Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server):W609–W612.
Taylor SA, Larson EL. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nat Ecol Evol*. 3(2):170–177.
Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* 9:322.
van der Ham R. 2015. On the history of the butternuts (*Juglans* section *Cardiocaryon*, Juglandaceae). *Palaeontogr Abteilung B* 293(1-6):125–147.
Wang H, Pan G, Ma QG, Zhang JP, Pei D. 2015. The genetic diversity and introgression of *Juglans regia* and *Juglans sigillata* in Tibet as revealed by SSR markers. *Tree Genet Genomes* 11:1.
Weber JA, Aldana R, Gallagher BD, Edwards JS. 2016. Sentieon DNA pipeline for variant detection—software-only solution, over 20× faster than GATK 3.3 with identical results. *PeerJ PrePrints* 4:e1672v1672.
Woeste K, Michler C. 2011. *Juglans*. In: Kole C, editor. Wild crop relatives: genomic and breeding resources: forest trees. Berlin/Heidelberg (Germany): Springer. p. 77–88.
Wolfe AD, Elisens WJ. 1995. Evidence of chloroplast capture and pollen-mediated gene flow in *Penstemon* Sect. *Peltanthera* (Scrophulariaceae). *Syst Bot*. 20(4):395–412.
Wu GA, Terol J, Ibanez V, Lopez-Garcia A, Perez-Roman E, Borreda C, Domingo C, Tadeo FR, Carbonell-Caballero J, Alonso R, et al. 2018. Genomics of the origin and evolution of *Citrus*. *Nature* 554(7692):311–316.
Xu Y, Zhang MY, Gao L. 2007. The research of interspecific hybridization in genus *Juglans*. *Decidous Fruits* 32:6–8.
Yu Y, Nakhleh L. 2015. A maximum pseudo-likelihood approach for phylogenetic networks. *BMC Genomics*. 16:S10.
Yuan X-Y, Sun Y-W, Bai X-R, Dang M, Feng X-J, Zulfiqar S, Zhao P. 2018. Population structure, genetic diversity, and gene introgression of two closely related walnuts (*Juglans regia* and *J. sigillata*) in Southwestern China revealed by EST-SSR markers. *Forests* 9(10):646.
Zhao P, Zhou HJ, Potter D, Hu YH, Feng XJ, Dang M, Feng L, Zulfiqar S, Liu WZ, Zhao GF, et al. 2018. Population genetics, phylogenomics and hybrid speciation of *Juglans* in China determined from whole chloroplast genomes, transcriptomes, and genotyping-by-sequencing (GBS). *Mol Phylogenet Evol*. 126:250–265.
Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, Weir BS. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28(24):3326–3328.