Phylogenetics and biogeography of Oleaceae

Julia Dupin, Cynthia Hong-Wa, Myriam Gaudeul, Guillaume Besnard

Phylogenetics and biogeography of the olive family (Oleaceae)

Julia Dupin[1*], Cynthia Hong-Wa[2], Myriam Gaudeul[3], Guillaume Besnard[1]

[1] CNRS, Université Paul Sabatier, IRD, UMR 5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 route de Narbonne, F-31062 Toulouse, France ; [2] Claude E. Phillips Herbarium, Delaware State University, 1200 N. DuPont Hwy, Dover, DE 19901, USA; [3] Institut de Systématique Evolution Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, Université des Antilles, 57 rue Cuvier, CP39, 75005 Paris, France

* Corresponding author: julia.dupin@gmail.com

ABSTRACT

Background and Aims: Progress in the systematic studies of the olive family (Oleaceae) during the last two decades provides the opportunity to update its backbone phylogeny and to investigate its historical biogeography. We additionally aimed to understand the factors underlying the disjunct distribution pattern between East Asia and both West Asia and Europe that is found more commonly in this family than in any other woody plant families.

Methods: Using a sampling of 298 species out of ca. 750, the largest in a phylogenetic study of Oleaceae thus far, and a set of 36 plastid and nuclear markers, we reconstructed and dated a new phylogenetic tree based on maximum likelihood and Bayesian methods and checked for any reticulation events. We also assessed the relative support of four competing hypotheses [Qinghai–Tibet Plateau uplift (QTP-only hypothesis), climatic fluctuations (Climate-only hypothesis), combined effects of QTP uplift and climate (QTP-Climate hypothesis), and no effects (Null hypothesis)] in explaining these disjunct distributions.

Key Results: We recovered all tribes and subtribes within Oleaceae as monophyletic, but uncertainty in the position of tribe Forsythieae remains. Based on this dataset, no reticulation event was detected. Our biogeographic analyses support the QTP-Climate hypothesis as the likely main explanation for the East-West Eurasian disjunctions in Oleaceae. Our results also show an earlier origin of Oleaceae at ca. 86 Mya and the role of Tropical Asia as a main source of species dispersals.

Conclusion: Our new family-wide and extensive phylogenetic tree highlights both the stable relationships within Oleaceae, including the polyphyly of the genus *Chionanthus*, and the need for further systematic studies within the family's largest and most under-sampled genera (*Chionanthus* and *Jasminum*). Increased sampling will also help to fine-tune biogeographic analyses across spatial scales and geological times.

KEYWORDS: Climatic fluctuations, *Chionanthus*, disjunction, divergence time, Eurasia, historical biogeography, *Jasminum*, Oleaceae, phylogenetics, Qinghai–Tibet Plateau, reticulation, Tropical Asia.

INTRODUCTION

Oleaceae are a plant family of ca. 750 species placed within 27 genera (Green, 2004; Banfi, 2014; de Juana Clavero, 2020; Li *et al*., 2020; Dupin *et al*., 2022; Hong-Wa *et al*., 2023). The family mostly includes woody plants, with many species of economic and ecological importance such as olives, ash trees, jasmines, forsythias, osmanthuses, privets and lilacs (Fig. 1). Current taxonomy recognizes five tribes (Fontanesieae, Forsythieae, Jasmineae, Myxopyreae, and Oleeae) within Oleaceae, and then four subtribes within Oleeae (Wallander and Albert, 2000). Some phylogenetic studies have further refined the infrafamilial classification with a backbone phylogeny that has Oleeae and Jasmineae, most likely, as sister tribes, with Forsythieae sister to them, then Fontanesieae sister to the clade formed by the three tribes, and lastly Myxopyreae as sister to all of them; although the positions of Fontanesieae and Forsythieae are not well supported compared to the other nodes (Dupin *et al.*, 2020; Dong *et al.*, 2022). Along with the updated tribal and subtribal circumscriptions and relationships, multiple phylogenetic studies at the generic level have presented much-needed revisions to the family's taxonomy. For instance, two new genera [*Chengiodendron* (Li *et al.*, 2020) and *Chrysojasminum* (Banfi, 2014)] were described while the genus *Tetrapilus* was resurrected (de Juana Clavero, 2020). In addition, all species belonging to the African *Chionanthus*, *Nestegis*, and *Comotranthus* were incorporated into *Noronhia*, *Notelaea*, and *Schrebera*, respectively (Hong-Wa and Besnard, 2013; Dupin *et al.*, 2022; Hong-Wa *et al*., 2023). The taxonomic changes of the last two decades, along with the new genetic data publicly available, advocate for a new phylogenetic tree of the family to serve as a framework for addressing questions on the evolution of the family's complex traits such as reproductive systems, ploidy numbers, and biogeographic history.

Based on current data on geographic distribution, Oleaceae species occur in all continents except Antarctica (POWO, 2023), as well as in all climate regions except the polar and subpolar zones. A few species, especially some members of the genera *Olea* and *Menodora*, are native to semi-desertic habitats (e.g., *O. europaea* subsp. *laperrinei*, Battandier and Trabut, 1911; *M. robusta*, Steyermark, 1932). However, species abundance is skewed towards East Asia (from East China and Japan to Southeast Asia), as about one-fourth of Oleaceae's species have part of their native distribution in (or restricted to) that region; mainly due to *Jasminum*, the most speciose genus in the family. Such abundance in East Asia contrasts with the small number of species that are native to the European continent (only ca. 3%), which nonetheless features a good number of charismatic and economically important species. These include the olive tree (*Olea europaea*), the common ash tree (*Fraxinus excelsior*), and the common lilac (*Syringa vulgaris*).

Oleaceae's subcosmopolitan distribution includes several examples of a rare geographic disjunction between East Asia and Europe; a pattern that is recorded in only nine families of woody plants. Indeed, out of the numerous woody plant genera in different families whose native distribution is restricted to the Eurasian continent, just 15 genera present a distribution with several species native to East Asia, and usually one (or very few) native to Europe, with a clear gap in the western portions of Asia (Browicz, 1992). Of these 15 genera, five belong to Oleaceae (Green, 1972; Browicz, 1992), three to Rosaceae, and the remaining seven families have only one example each (Browicz, 1992). Within Oleaceae, one species in the genera *Ligustrum* (*L. vulgare*), *Fontanesia* (*F. phillyraeoides*), *Forsythia* (*F. europaea*), and *Osmanthus* (*O. decorus*), and two non-sister *Syringa* species (*S. josikaea* and *S. vulgaris*) are native to Europe, while all their congeners are found in East and South Asia.

Two main explanations for the disjunctions in plant distributions between East Asia and both West Asia and Europe are (i) the uplift of the Qinghai–Tibet Plateau (QTP), with a likely start in the Eocene and a subsequent important phase of rapid uplift in the Miocene, along with the onset of aridification in Central Asia during the Miocene, and (ii) the climatic fluctuations during the Pliocene and Quaternary. The uplift of the QTP (Harrison *et al*., 1992) created a geographical barrier for plants, pollinators, and dispersers, thereby leading to a geographic isolation among them. That uplift could also be linked to the climatic changes in the region (Miao, 2012; Lu *et al*., 2020) that contributed to the aridification of parts of Asia. Arid and semi-arid climates and a high level of xerophytic species endemism mark the disjunction zone between East and West Asia, which corresponds to most of today's Irano-Turanian region (Djamali *et al*., 2012; Manafzadeh *et al*., 2014, 2017). This contrast between xerophytic aspects in the Irano-Turanian region and mesophytic conditions on both its eastern and western sides is considered a contributor to such disjunction (Browicz, 1992). Extinctions due to the uplift of the QTP are presented as the likely explanation for disjunctions in several contemporaneous plant groups, such as Ranunculaceae, Rosaceae, and Pinaceae (Sun *et al*., 2001; Sun, 2002; Zhang *et al*., 2006; Qiao *et al*., 2007). For groups that originated much later, the climatic fluctuations that started during the Pliocene might be a more fitting explanation. The Pliocene included not only a transition from a relatively warm climate to the cooler climate of the Pleistocene, but also the establishment of Northern Hemisphere ice sheets and the beginning of glacial-interglacial cycles (Webb and Bartlein, 1992; Zachos *et al*., 2001; John and Krissek, 2002). With favorable climate conditions appearing and disappearing cyclically, the fragmentation of species distribution was likely frequent too (e.g., Fiz-Palacios *et al*., 2010; Tu *et al*., 2010). Finally, for some clades, it is also possible that both events in combination are responsible for their distribution.

In this study, we use a new phylogeny of Oleaceae as a framework to test biogeographic hypotheses regarding its history. Using an extensive and representative sampling (especially for *Jasminum*, the largest genus), and incorporating the latest fossil data in the family, we estimated a dated tree for Oleaceae. In this framework of divergence times, we aimed to answer general questions about the family's biogeographic history: (1) what is the family's center of origin, and does it correspond to its current center of diversity (East Asia)?; (2) what are the general patterns in the directionality of dispersal events?; and (3) when did the East-West Eurasian disjunctions happen, and how common was an ancestral widespread distribution between East Asia and Europe? Finally, to expand on question (3), we tested if those disjunctions we see today were a result of the uplift of the QTP (QTP-only hypothesis), a result of the climatic fluctuations during the Pliocene and Quaternary (Climate-only hypothesis), a result of both the QTP uplift and the climate (QTP-Climate hypothesis), or none of them (Null hypothesis).

MATERIALS AND METHODS

Plant sampling and DNA sequencing

Our final sample list included 402 accessions representing 300 species (298 species of Oleaceae and two species of Carlemanniaceae as outgroups; Supplementary Table S1). This dataset included 126 accessions newly sequenced for this study and 276 accessions whose data were publicly available on GenBank or the European Nucleotide Archive databases. Among species newly sequenced, we primarily sampled herbarium specimens of all major Oleaceae lineages, including representatives over different biogeographic areas.

Of the 126 newly sequenced accessions for this study, 110 accessions were sampled from herbarium specimens, seven accessions from living collections at CEFE Montpellier or field collections, and nine accessions were obtained from KEW DNA bank. For the 110 accessions from herbaria, and the seven from living collections or field-collected ones, we extracted total genomic DNA from ca. 10 mg of dried leaves. We ground the samples in 2-mL tubes with metal beads using a TissueLyser (Qiagen Inc., Texas). We then extracted the DNA following the BioSprint 15 DNA Plant Kit protocol (Qiagen Inc.) and eluted the extracted DNA in 200 μL of AE buffer. For all 126 samples, we used a genome skimming approach to sequence nuclear and plastid genomes. We constructed the sequencing libraries using 10 to 200 ng of double-stranded DNA with TruSeq Nano HT Sample kit (Illumina), following default instructions. Each sample was paired-end sequenced (150 bp) on an Illumina HiSeq 3000 lane, multiplexed, and shotgun sequenced at the Genopole platform of Toulouse.

To assemble the different regions used here [the plastid genome, the nuclear ribosomal cluster (nrDNA), and 17 low-copy nuclear markers], we mapped short reads to a bait reference sequence using GENEIOUS v9.0.5 (https://www.geneious.com), NOVOPlasty (Dierckxsens *et al.*, 2017), or aTRAM (Allen *et al.*, 2018). All three software use an automated iteration approach, where reads are first mapped to a reference, and, once a contig is assembled, it then serves as the query for the next iteration and so on, up to the maximum number of user-specified iterations. To assemble the plastid genome and the nrDNA cluster, we used for reference a sequence from a closely related species that was already available. We annotated all plastid genomes using GeSeq (Tillich *et al*., 2017) and checked the quality of the annotations in GENEIOUS.

Accessions and phylogenetic markers selection

We then constructed a sequence data matrix by adding data from public databases to those newly generated here, but we used a few criteria to retain an accession in our dataset. The first step of the pipeline was to remove any accessions with only one genetic marker available. The second step was to remove accessions that, once we quickly estimated a Maximum Likelihood (ML) tree, did not group with others from the same genus or species, and for which we could not easily ascertain their identification. We associate such discrepancies in the topological placement of accessions (from GenBank) with identification issues, or the low number and/or informativeness of sequences available for a given accession. For instance, accessions that only had sequences for a few plastid genes (i.e., very conserved sequences), and no sequences for intergenic plastid spacers or nuclear markers, tended to be placed erroneously in the tree. After a few rounds of data cleaning, we retained 402 accessions (Supplementary Table S1).

Our matrix included a combination of 18 plastid markers, the nrDNA cluster, and 17 low-copy nuclear markers originating from the Angiosperms353 list of single-copy protein-coding genes (Johnson *et al.*, 2019). For plastid markers, we used the most common ones in the dataset to maximize the number of shared markers between accessions, and added phylogenetically-informative ones. The final list included ten genes (*matK*, *ndhF*, *rbcL*, *rpoB*, *rpoC1*, *rpoC2*, *rps16*, *trnK*, *ycf1*, *ycf2*) and eight intergenic spacers (*atpB-rbcL*, *psbA-trnH*, *psbJ-petA*, *rpl32-trnL*, *trnL-trnF*, *trnQ-rps16*, *trnS-trnG*, *trnT-trnL*). For nuclear markers, we included nrDNA sequences because it is the most common nuclear marker available for plants. We transformed the nrDNA sequences from regular nucleotide coding to purine-pyrimidine-only coding (usually referred to as RY coding). It has been shown that the GC content of nrDNA sequences in Oleaceae (especially

in the tribe Oleeae) can vary a lot, and species with high levels of GC tend to be erroneously grouped in phylogenetic analyses (Hong-Wa and Besnard, 2013; Ha *et al.*, 2018; Olofsson *et al.*, 2019). With RY coding, one can effectively reduce the influence of biased GC content (Phillips *et al.*, 2004). Finally, for a selection of 20 species chosen to represent tribes and subtribes in Oleaceae and the outgroup [18 Oleaceae and two Carlemanniaceae, for which we either generated short-read data in moderate to high depth (> 20×) or used data available freely], we assembled partial sequences for 17 low-copy nuclear markers from the list in the Angiosperms353 set. These markers have the following gene identifiers in the *Olea europaea* genome v1.0 available through Phytozome (Goodstein *et al.*, 2012): Oeu006114.1, Oeu006265.1, Oeu006511.1, Oeu007384.1, Oeu015466.1, Oeu016021.3, Oeu018164.2, Oeu021892.1, Oeu022026.1, Oeu027950.1, Oeu031929.1, Oeu045518.2, Oeu046246.1, Oeu053181.2, Oeu056039.1, Oeu056679.1, and Oeu062177.2.

Tree search

We uploaded a concatenated matrix of all regions (18 plastid regions, nrDNA, and 17 low-copy nuclear markers) into the program IQ-Tree v2.0.6 (Minh *et al.*, 2020a) to estimate an ML species tree and assess the genealogical concordance between partitions. To estimate the tree, we first did ten rounds of substitution model and partition scheme selection using ModelFinder (Kalyaanamoorthy *et al.*, 2017). We then used the best partition scheme (the one with the lowest Bayesian Information Criterion [BIC] score; Supplementary Table S2), along with a concatenation approach with an edge-linked proportional partition model (Chernomor *et al.*, 2016), and estimated a species tree. We assessed node support through 1000 ultrafast bootstrap (UFBootstrap) replicates (Hoang *et al.*, 2018) and 1000 replicates for the non-parametric SH-aLRT topological test (Guindon *et al.*, 2010).

Finally, we calculated the gene concordance factor (gCF) and the site concordance factor (sCF) for each branch of the species tree (Minh *et al.*, 2020b). With gCF, we represented the proportion of gene trees concordant with a given branch and, with sCF, the fraction of alignment sites supporting such branches. In total, we had 14 regions to estimate gCF values. These 14 regions were the ones resulting from the best partition scheme (Supplementary Table S2), where initial markers were combined into larger regions based on their estimated models of sequence evolution, and included five nuclear regions and nine plastid ones. In the case of sCFs, values were calculated based on a dataset of 120,057 sites (56% nuclear, 44% plastid).

Network analyses

Using a subset list of species (i.e., the 18 Oleaceae species representing major lineages for which we assembled the 17 low-copy nuclear markers; Supplementary Table S2), we evaluated the presence and location of reticulation events in Oleaceae's history. We used PhyloNetworks v0.14.3 (Solís-Lemus and Ané, 2016; Solís-Lemus *et al.*, 2017), PhyloPlots v0.3.1 (implemented in PhyloNetworks; Solís-Lemus *et al.*, 2017), and QuartetNetworkGoodnessFit v0.3.3 (Cai and Ané, 2021) packages in Julia (Bezanson *et al.*, 2017) to infer phylogenetic networks from gene tree quartets. In PhyloNetworks, we input individual loci trees for each partition (for all plastid and nuclear regions; trees estimated in IQ-Tree), and these trees were summarized into a table of quartet concordance factors. This table shows the frequency of all subsets of four species. Using our

species tree in combination with the table of quartets, we tested four scenarios: one where there are no reticulation events (i.e., the species tree is the most likely representation of the family's evolutionary history; net0), and three other scenarios where the number of reticulation events allowed is increased by one up to three total events (net1, net2, net3, respectively). We simulated each of those scenarios ten times. Finally, to test for the likelihood of a scenario given the data, we used the package QuartetNetworkGoodnessFit and ran 1000 simulations each time.

Divergence time estimation

We dated the Oleaceae tree using the Bayesian approach in BEAST2 v2.6.7 (Bouckaert *et al*., 2019). The dataset was reduced from the 402 accessions to 285 accessions, representing 275 species. We decided to reduce the original dataset to allow for less computational time. This reduction was done in two steps: first, by reducing the number of accessions (numbers mentioned previously) but keeping the proportions in number of species per genus; second, by decreasing the number of characters in the DNA matrix from 120,058 to 33,808 (keeping eight nuclear regions with 24,456 characters, and seven plastid regions with 9,352 characters. Each of the partitions had its substitution model estimated separately (i.e., un-linked models), with initial values for site models the same for all partitions: gamma site model, gamma category count set to four, substitution model set as GTR, and all other parameters left as default values. Regarding clock models, we linked models for all plastid partitions, and left models for nuclear regions unlinked. We chose an Optimized Relaxed Clock (ORC) prior for all clock models. Finally, we fixed the topology of the tree throughout the runs using as input the topology of the ML tree from the IQ-Tree analyses. The topology was fixed by setting up the starting tree, and by reducing the weight of parameters "wide exchange", "narrow exchange", "Wilson-Balding", and "subtree-slide" to zero as recommended by BEAST2 tutorials on topology fixing.

We used a combination of six primary calibration points and three secondary points for the estimation of divergence times. According to the results of Palamarev (1989), we added minimum stem ages for *Chrysojasminum fruticans* [2.58 million of years (My)], *Olea* subgenus *Olea* (23.03 My), and *Phillyrea* (5.33 My). Based on the findings of Mathewes *et al*. (2021), we added a minimum stem age for Fraxininae of 47.8 My. Additionally, given the findings of Wu et *al*. (2021), we set the minimum stem age of both *Fraxinus* sections *Dipetalae* and *Ornus* as 32.1 My. Regarding secondary calibration points, we set the minimum crown ages of the *Olea europaea* complex (3.5 My), and *Olea* subgenus *Olea* (13.8 My) based on results of Besnard et *al*. (2009), and we set the minimum and maximum stem ages for the whole Oleaceae family (59.07 My and 87.2 My, respectively) based on the 95% higher density probability (HPD) boundaries of the stem age of Oleaceae from Magallon et *al*. (2015)'s UCLN analyses.

For all calibration points except the one for Oleaceae stem age, we specified a log normal prior, where the offset values were the minimum ages. We set mean values as 2.5 to the *Chrysojasminum fruticans* and *Phillyrea* calibration points, and all others had mean values set as 1.0. For variance levels, we set those of *C. fruticans* and *Phillyrea* calibration points as 0.75, and all other points had their variance set to 1.0. We defined those mean and variance values based on the results of a preliminary run. Finally, for the Oleaceae stem age, we specified a uniform distribution prior with an offset value set to zero, and lower and upper limits values that match minimum and maximum ages. In total, we ran two chains of 400 million generations each.

Historical biogeography

We used a pruned version of the dated ML tree to better represent the distribution of species, based on the distribution of all accepted species today (POWO, 2023; Table 1). The final tree (the biogeographic tree) included 211 Oleaceae species. For these analyses, we classified native species distributions into seven areas: (A) Temperate Asia, (B) Tropical Asia, (C) Australasia, (D) Europe, (E) Africa, (F) Northern America, and (G) Southern America (Supplementary Table S1).

We used the R package BioGeoBEARS v1.1.2 (Matzke, 2018a, b; Matzke *et al.*, 2019) to estimate ancestral ranges within Oleaceae. BioGeoBEARS implements ML methods that replicate the key assumptions of DEC (dispersal–extinction–cladogenesis; Ree and Smith, 2008), DIVA (dispersal–vicariance analysis; Ronquist, 1997), and BayArea (Bayesian Inference of Historical Biogeography for Discrete Areas; Landis *et al.*, 2013). Collectively, these methods allow for processes of within-area speciation, vicariance, range expansion, range contraction, and founder-event speciation (range switching at nodes; Matzke, 2014). We implemented six methods per hypothesis: DEC, DIVALIKE, and BAYAREALIKE, and their variation with the addition of the jump dispersal parameter (*j*). For all of those, we also included the parameter *w*, a matrix exponent that we set to be freely estimated to optimize dispersal matrices. The *maxareas* parameter was set to four.

Finally, we incorporated time-stratified dispersal multiplier matrices into all models to account for geologic and climatic changes over time. We set out a total of five time periods: from Oleaceae's origin (ca. 86 My) up to 65 My, from 65 to 33.9 My, from 33.9 to 20 My, from 20 to 5.3 My, and from 5.3 to today. The dates 65, 33.9, 20, and 5.3 My mark, respectively, the Cretaceous–Paleogene boundary (Renne *et al.*, 2013), the Eocene/Oligocene transition (Cohen *et al*., 2022), the rapid uplift phase of the QTP (Harrison *et al*., 1992), and the beginning of the Pliocene (that led into the Pleistocene; Cohen *et al*., 2022). To test the influence of the QTP uplift and the climate during the Pliocene (and Pleistocene) on the range evolution between Asia and Europe, we set up four sets of dispersal matrices (all with the same time periods) to correspond to the four hypotheses being tested: i) a first matrix for the Null hypothesis, where range evolution is only dependent on the distances between the regions, ii) a second set of matrices to simulate the influence of the QTP uplift only (QTP-only hypothesis), iii) a third one to, in turn, simulate the influence of the Pliocene climate only (Climate-only hypothesis), and iv) a final one where the ranges evolved in a setup where both the QTP uplift and the climate influence them (QTP-Climate hypothesis; Supplementary Methods S1). Values in the dispersal matrices were based on the dispersal graphs presented by Landis (2017), where areas connected by black lines had a dispersal probability between them of 0.7, those connected by grey lines only or a combination of black and grey lines had their dispersal probability defined as 0.3, and the dispersal between areas without a connecting line in the dispersal graph was set as 0.1. Additionally, the combined effects of the QTP uplift and the climatic fluctuations on the dispersal between some areas were represented by a value of 0.01. We compared all 24 models (four hypotheses, six models per hypothesis) in parallel, and used the Akaike Information Criterion (AIC) and Akaike weights (Burnham and Anderson, 2002) to rank the fit of each model to the data.

Adopting the best model, we then ran a Biogeographic Stochastic Mapping (BSM) analysis (Dupin *et al.*, 2017). We estimated event frequencies by taking the mean and standard deviation of event counts from 100 BSMs. We specifically looked at the frequency of vicariance, founder, and

extinction events between Asia and Europe, and their timing, to contextualize the current examples of East-West Eurasian disjunctions.

RESULTS

DNA markers

Our final character matrix had 402 accessions with 14 partitions and 120,057 total sites (with ca. 55% of missing data), where about 40% of the total sites presented unique patterns, and ca. 20% were parsimony-informative (Supplementary Table S2). Further, this dataset showed 67,154 nuclear-derived sites (56% of total sites) and 52,903 plastid-derived sites (44%). We also had similar proportions between nuclear and plastid sites for the number of parsimony-informative sites. However, most of the missing data were in the nuclear sequences. Following the selection of substitution models in ModelFinder, the original 36 partitions were rearranged in a new, 14-partition scheme (Supplementary Table S2).

Tree search

We found that all tribes and subtribes were monophyletic, receiving maximum support from both the UFBootstrap and SH-aLRT test results. Based on this dataset, the most likely topology among tribes was (Myxopyreae, (Fontanesieae, (Forsythieae, (Jasmineae, Oleeae)))) (Fig. 2) and the subtribes of Oleeae were represented as (Schreberinae, (Ligustrinae, (Fraxininae, Oleinae))) (Fig. 2, Supplementary Figures S1 and S2). This backbone topology of the family was highly supported based on UFBootstrap and SH-aLRT test (Table 2). The only exception was the clade (Forsythieae, (Jasmineae, Oleeae)), with values of only 93 for UFBootstrap support and 19.4 for the SH-aLRT test.

Values for concordance factors (gCF and sCF) for tribes and subtribes, and clades between them, varied but in a correlated way (e.g., a high gCF value coupled with a high sCF value), and were moderate to high (between ca. 50 and 100; Table 2, Supplementary Figure S2). There were two exceptions, the first one was the (Forsythieae, (Jasmineae, Oleeae)) clade, with a gCF value of 21.4 and sCF of 27.1. The second one was the (Ligustrinae, (Fraxininae, Oleinae)) clade that showed a moderate gCF of 50.0 but a low sCF of 21.6.

Beyond the phylogeny backbone, we found several examples of incongruences between the current taxonomy and the placement of accessions in our ML tree (Supplementary Figures S1 and S2), notably in *Chionanthus*. Species of *Haenianthus* (Caribbean) were shown as sister to most *Chionanthus* species from the Caribbean, Central and South America (*C. compactus, C. domingensis, C. implicatus, C. filiformis, C. fluminensis, C. trichotomus, C. jamaicensis, C. panamensis,* and *C. pubescens*). *Chionanthus pygmaeus* and *C. virginicus*, which are native to North America (specifically, USA and Mexico), grouped with species of *Cartrema* (N America) and *Chengiodendron* (SE Asia). *Chionanthus ligustrinus* (Caribbean) is sister to *Priogymnanthus* species (S America), and *C. mala-elengi* and *C. parkinsonii* (Thailand), along with one Chinese species of *Tetrapilus* (*T. caudatilimbus*), formed a clade sister to *Noronhia* (Africa). Another species of *Tetrapilus*, *T. rubrovenius* (Indonesia), also had its placement just outside of the genus; it formed the sister clade to *Tetrapilus* along with *Chionanthus polygamous* and *C. rupicolus* (SE

Asia). Additionally, *Ligustrum* species formed a clade nested within *Syringa* (with *S. reticulata* as the sister lineage to *Ligustrum*); *Picconia* (Macaronesia) was shown nested within *Phillyrea* (Mediterranean).

Network analysis

Out of the four reticulation scenarios that we tested (i.e., no reticulation event, one, two, or up to three reticulation events allowed), the most supported one was the no reticulation event. The negative log pseudolikelihood scores of the different scenarios were the following: -Ploglik = 6451.14 [net0], 5534.63 [net1], 5453.90 [net2], and 5418.53 [net3]. Here, there was a sharp improvement in likelihood between net0 and net1, and the following values increased more gradually thereafter. Further, the results from the Goodness-of-Fit test showed that, when net0 and net1 were tested against the data, net0 already fit the data adequately (p-value = 0.9). In other words, our ML tree was a reliable representation of Oleaceae's evolutionary history, based on this dataset.

Divergence times

Our analyses showed that Oleaceae originated around 86 million years ago (Mya; 95% HPD: 83.7-87.2 Mya), placing its origin during the Santonian age of the Upper Cretaceous (Table 3; Fig. 3). The diversification of the family's tribes is estimated to have taken place mainly after the Cretaceous–Paleogene (K–Pg) event, with Myxopyreae being the oldest tribe [ca. 68 Mya (60.2-75.3 Mya)] and Fontanesieae as, potentially, the youngest one [crown age of 0.3 Mya (0-1.5 Mya), but with a very long stem branch spanning almost 80 My]. The subtribes of tribe Oleeae all started diversifying between the Late Eocene and Early Oligocene, with Fraxininae as the oldest one [41.0 Mya (37.1-45.5 Mya)] and Ligustrinae the youngest [32.5 Mya (23.6-42.9 Mya)].

Historical biogeography and biogeographic stochastic mapping

Our results showed that the dispersal matrices representing the combined influence of the QTP uplift, with a likely onset in the Eocene followed by a relevant phase of rapid uplift in the Miocene, and the climatic fluctuations during the Pliocene and Quaternary (QTP-Climate hypothesis) fit the data better than any of the other remaining hypotheses tested. Specifically, the best biogeographic model (out of the 24 compared) is the DEC+*j* with a log-likelihood score of -357.5, an AIC value of 723.0, and an Akaike weight of 0. 99 (Supplementary Table S3). Under this biogeographic model (QTP-Climate DEC+*j* model), the family's early biogeographic history likely took place in Asia, but with varying degrees of certainty for the estimated ranges. The family's ancestral range is uncertain, with about 47% probability that the ancestral range included Tropical Asia (area B, Fig. 3). For the tribes, the ancestral state is likely Tropical Asia for Myxopyreae (76% likelihood), Jasmineae (61%), and Oleeae (62%). For Fontanesieae, the estimated range was a combination of Tropical Asia, Temperate Asia, and Europe (97%); and it was most likely Temperate Asia (90%) for Forsythieae (Fig. 3, Supplementary Figure S3).

Most of the dispersal events throughout Oleaceae's history had as their source Tropical and Temperate Asia, the two regions that currently harbor most of the species in the family (Fig. 4,

Supplementary Table S4). Out of the total mean number (88) of dispersal events estimated, 45 of them were from Tropical Asia towards other areas; the second highest source of dispersals being Temperate Asia. Specifically, regarding the events originating in Temperate Asia, the majority of them (15 out of 24) represented range expansions into Tropical Asia. Temperate Asia not only had most of its dispersals towards Tropical Asia, it also served as the sink of almost half of the dispersals from Tropical Asia, making its species diversity heavily linked to that of Tropical Asia. The results also showed specific patterns in the balance of outgoing and incoming dispersal events for the given areas. While Tropical Asia served twice as more as a source of dispersal events than it did as a sink (45 out vs. 21 in), Temperate Asia and Northern America had well balanced proportion of events, and the remaining of the areas all had at least twice as many incoming events as there were outgoing ones (i.e., the opposite pattern seen for Tropical Asia).

When looking at the timing of the events that likely contributed to the disjunctions between Asia and Europe (extinction, vicariance, or founder events), we estimated that most of those events likely happened in the last 20 My (Fig. 5). Our results also show a sharp increase in the number of those events after 20 Mya (Fig. 5), which is well aligned with what our best biogeographic model represents: starting around 20 Mya, the combined effects of the QTP uplift and the climatic fluctuations during the Pliocene and Quaternary. Lastly, we found that ranges that included at least one area in Asia and Europe increased gradually towards its peak between 20 and 5.3 Mya, with a mean of ten ranges at that time, decreasing afterwards (Supplementary Figure S4).

DISCUSSION

Phylogeny

This study is the first to present a large-scale phylogeny of the Oleaceae with 402 accessions representing 298 species, ca. 40% of the family's diversity. Using an extensive set of phylogenetic markers, it offers a fresh outlook on relationships as much among the major lineages (tribes and subtribes) as within them. All tribes and subtribes were recovered as monophyletic (Fig. 2, Supplementary Figures S1 and S2), supporting previous studies that used fewer and/or different markers (Wallander and Albert, 2000; Lee *et al*., 2007; Olofsson *et al*., 2019; Dupin *et al*., 2020; Dong *et al*., 2022). The recovered topology has the tribe Myxopyreae as sister to a clade formed by the four other tribes. Within that large clade, tribes Jasmineae and Oleeae are sister groups, being then sister to tribe Forsythieae, and this clade of three tribes is then sister to tribe Fontanesieae (Fig. 2, Supplementary Figures S1 and S2). Among the subtribes of Oleeae, subtribe Schreberinae is sister to all other subtribes, of which subtribe Ligustrinae is then sister to the clade formed by the subtribes Fraxininae and Oleinae (Fig. 2, Supplementary Figures S1 and S2). As discussed in Wallander and Albert (2000), the tribes and subtribes are recognized by distinctive synapomorphies, including anatomical, chromosomal, and morphological features. Representative members of each tribe and subtribe are shown in Fig. 1. Our results further showed that relationships among tribes and subtribes had high branch support with one exception, the placement of Forsythieae. Indeed, tribe Forsythieae is confidently more closely related to Oleeae and Jasmineae than it is to other tribes, but it was not possible, with this dataset, to pinpoint its exact placement. Low support for the placement of this clade was shown in other studies as well (Wallander and Albert, 2000; Dupin *et al.*, 2020; Dong *et al.*, 2022). Further, tribe Oleeae likely resulted from a hybridization event between the ancestral lineage of Forsythieae and another lineage (either Jasmineae ancestral lineage or an extinct one; Dong *et al*., 2022). In addition, whole

genome duplication events are suspected in Forsythieae's ancestral lineage (Pei *et al.*, 2024). The likely high levels of genome exchanges between ancient lineages may thus explain the lower level of support between these three tribes.

Except for tribe Oleeae, all other tribes include only a few genera, whose relationships are well supported (Fig. 2, Supplementary Figures S1 and S2). For instance, within Myxopyreae, the genera *Dimetra* and *Nyctanthes* form a clade sister to *Myxopyrum*, whereas within Jasmineae, *Chrysojasminum* is sister to the clade formed by *Jasminum* and *Menodora*. It is worth noting that *Jasminum* section *Primulina* is sister to the clade formed by all other sections (Supplementary Figures S1 and S2) and is distinct from all other members of *Jasminum* by having yellow flowers (vs. white flowers) ~~and may represent a distinct genus~~. The tribes Fontanesieae and Forsythieae include only one and two genera, respectively. Within tribe Oleeae, all subtribes also include only one or two genera at the exception of subtribe Oleinae where the bulk of the genera in the family Oleaceae as a whole is found (currently 14 described genera).

Within subtribes, all genera were also monophyletic except for *Ligustrum*+*Syringa*, with a paraphyletic *Syringa*; *Phillyrea*+*Picconia*, with a paraphyletic *Phillyrea*; and a pervasively polyphyletic *Chionanthus*; and lastly a polyphyletic *Tetrapilus* (Fig. 2, Supplementary Figures S1 and S2). These intergeneric patterns have also been previously recovered in other, sometimes smaller, studies (Li *et al.*, 2002; Li *et al.*, 2012; Hong-Wa and Besnard, 2013; Olofsson *et al*., 2019; Dupin *et al.*, 2020; Dong *et al*., 2022). The convergence of these multiple studies to the same results calls for further lower-level phylogenetic assessments, with exhaustive sampling, to help decipher generic relationships and improve the infrafamilial taxonomy of the Oleaceae. Overall, intergeneric relationships within subtribe Oleinae remain unclear, but we have a clearer picture of the extent of the non-monophyletic patterns. For instance, members of the genera *Chionanthus* and *Tetrapilus* are found in nine and three distant positions, respectively (Fig. 2, Supplementary Figures S1 and S2). There is also a strong geographic signal wherein members from the same geographic area cluster together on the phylogenetic tree regardless of taxonomy. Such geographic patterns have also been previously recovered in other studies (e.g., Hong-Wa and Besnard, 2013; Olofsson *et al*., 2019), suggesting perhaps a morphological disparity at smaller spatial scales that is correlated with environmental factors, which has consequently biased morphologically-based taxonomy. The new topology provides a robust framework to guide future generic realignments within subtribe Oleinae and to investigate synapomorphies that unite sister taxa described under different generic names. Moreover, at the infrageneric levels, relationships were largely characterized by short branches in the tree, with levels of support varying from high to low (Supplementary Figures S1 and S2). A further understanding of infrageneric relationships may require a population genetics approach combined with the analyses of other types of data such as morphological, anatomical, and phytochemical.

The two largest genera of Oleaceae, namely *Chionanthus* and *Jasminum*, were represented in this study by 26 and 51 species, respectively. With about a fourth of its estimated diversity (POWO, 2023) included herein, the largest sampling in a phylogenetic study to date, *Jasminum* was recovered as monophyletic. While this pattern may change if all *Jasminum* species were to be sampled, the narrowly-defined *Jasminum*, after the exclusion of *Chrysojasminum* [characterized by alternate leaves and yellow flowers (Banfi, 2014)] and *Jasminum* section *Primulina* [characterized by opposite leaves and yellow flowers, and gapped testa (Rohwer, 1997; Green, 2001), and herein sister to all other *Jasminum* (Supplementary Figure S1)], is united by synapomorphic conditions such as never-yellow flowers and sarcotesta (Rohwer, 1997; Banfi,

2014). By contrast, *Chionanthus*, which was represented in this study by about a fifth of its diversity (POWO, 2023), showed up in nine separate positions along the phylogenetic tree (Fig. 2, Supplementary Figures S1 and S2). Obviously, this genus represents several different entities, whose inclusion within *Chionanthus* may have resulted from consideration of plesiomorphic and/or homoplastic features. In fact, members of *Chionanthus* had initially been recognized in one of at least six genera (Stearn, 1980; Hong-Wa and Besnard, 2013), which were subsequently subsumed under *Chionanthus* (Stearn, 1976, 1980). The delineation of these distinct genera was thought to be arbitrary due to overlapping morphological characters that did not justify recognizing several entities separate from *Chionanthus* (Johnson, 1957; Stearn, 1976). Yet, the largely defined *Chionanthus*, mainly recognized based on the shape, size, and degree of fusion of the corolla lobes, also lacks any clear synapomorphy, with wood anatomy, chemotaxonomy, and molecular data suggesting instead several recognizable entities (Harborne and Green, 1980; Baas *et al*., 1988; Hong-Wa and Besnard, 2013; Olofsson *et al*., 2019; Dupin *et al*., 2020). This highlights the incongruence between different types of data in delimiting this genus. Moreover, the strongly-supported, multiple phylogenetic placements of *Chionanthus* highlight the challenges of morphology-based taxonomy and warrant a recircumscription of this genus. However, the different clades of *Chionanthus* do not necessarily correspond to the subsumed genera within which species were originally described, requiring further species-level taxonomic work.

Divergence times

The family's estimated age of ca. 86 My and that of its tribes are much older than what was reported in other studies. Dong *et al*. (2022) reported an age for the crown node of Oleaceae as ca. 60.51 My. Olofsson *et al*. (2019) reported that the split between tribes Forsythieae and Oleeae happened at ca. 60 Mya, while we found a mean crown age for the clade Forsythieae+Oleeae+Jasmineae of 78 My. We attribute these differences to the set of calibrations used on those studies and this one. In contrast to the set of four and five calibrations points used on those studies, respectively, here we used a total of nine points, including three recently described *Fraxinus* fossils (Mathewes *et al*., 2021; Wu et *al*., 2021). Specifically, the placement of primary calibrations in four genera (i.e., *Chrysojasminum*, *Fraxinus*, *Olea*, *Phillyrea*) gives these results a broader span of reference points for divergence times.

Historical biogeography

The ancestral range of Oleaceae. Our estimation of the historical biogeography of Oleaceae shows that the first half of its diversification likely happened in Asia; although this early episode remains difficult to disentangle. While our results point to an Asian center of origin, their ambiguity indicates that the complexity of the directionality of dispersals (e.g., high frequency of dispersal events between Tropical and Temperate Asia areas) and the current distribution of species likely require a larger dataset than the one used for the present work. Indeed, in our final dataset for the biogeographic analyses, we pruned down our sampling to a little less than 30% of the species. A more extensive dataset covering at least 50% of the species, while maintaining our approach to balance out clades and areas represented, might do a better job at tackling some basic questions on the historical biogeography of this family.

This study also shows that for questions regarding origin of tribes, subtribes, or other smaller clades, a more extensive sampling of said clades is required. Even though we have spanned our sampling over all tribes in the family, we generally found less resolution for estimated ancestral ranges at the family level than studies that focused on subgroups of Oleaceae such as genera [e.g., *Fraxinus* and *Notelaea* (Hinsinger *et al*., 2013; Dupin *et al*., 2022)], subtribes [e.g., Schreberinae (Hong-Wa *et al*., 2023)], or tribes [e.g., Forsythieae (Ha *et al*., 2018)], all of which were proportionally sampled more extensively in each study.

Directionality of dispersal events. In Oleaceae, the strong directionality of dispersal events from Tropical and Temperate Asia areas towards other areas (Supplementary Table S4) may result from the age of Asian lineages. Tropical and Temperate Asia areas hold about 386 extant species that have their partial or total native range in those areas. This is equivalent to a little more than 50% of the accepted species today. Our ancestral range estimation showed that roughly the first half of the family's diversification likely happened in Asia, making this region the main source of migrants for the past 80 My.

The high number of dispersal events between Tropical and Temperate Asia indicates that even though there were periods where climatic barriers reduced dispersals between these areas, the exchanges between them remained significant over time. Geologic data shows an aridity belt existed from the western-most part of China to the eastern coast for most of the Paleogene (65 to 28 Mya) and only decreased towards the Early Miocene (ca. 23 Mya; Tiffney and Manchester, 2001; Guo *et al*., 2008). This belt, after receding, redeveloped towards the Late Miocene (ca. 7 Mya), limiting once again dispersals between northern and southern parts of China (Guo *et al*., 2008).

The uplift of the QTP, climatic variations in the Pliocene and Pleistocene, and the East-West Eurasian disjunctions. After testing different biogeographic models against our data, we found that the combined effects of the uplift of the QTP during the Eocene and Miocene and those of the climatic variations during the Pliocene and Pleistocene are likely the main causes of East-West Eurasian disjunctions in Oleaceae. Other studies on clades with this type of disjunction also support that vicariance led to this type of distribution. In *Paliurus* (Rhamnaceae), the divergence time between the East Asian and Mediterranean species was estimated as 18 Mya, a timing that overlaps with the QTP uplift (Chen *et al*., 2017). The fossil record of *Hedera* (Araliaceae), with multiple accessions from Europe and Asia, points to *Hedera* species as "Tertiary relicts" because they show widespread occurrence in the northern hemisphere during warm periods of the Miocene and Early Pliocene that became geographically restricted during cooler periods of the Late Pliocene and Pleistocene (Green *et al*., 2011). In *Cedrus* (Pinaceae), the present distribution of its species in several isolated regions suggests that populations got broken down due to vicariance events during climatic oscillation in the Tertiary; an explanation that is applicable to the divergence between the Himalaya cedar and its sister clade comprising Mediterranean cedars (Qiao *et al*., 2007).

CONCLUSION

We presented here a new phylogenetic tree for the Oleaceae with an extensive sampling and a large selection of genetic markers. With these data, we recovered all tribes and subtribes as monophyletic, as well as almost all genera, the exception being the paraphyletic *Phillyrea* and *Syringa*, and the polyphyletic *Chionanthus* and *Tetrapilus*. The dated version of this tree has

divergence times older than what has been found in other studies, placing the early diversification of the Oleaceae in the Upper Cretaceous (ca. 86 Mya). These older dates are expected given the larger set of fossil calibrations used here, including recently described fossils in *Fraxinus*.

We used this phylogenetic framework to answer questions about the biogeographic history of Oleaceae, including a peculiar disjunction observed in five of its genera (East-West Eurasian disjunctions). However, the early part of the biogeographic history of Oleaceae remains difficult to disentangle, with an ancestral area that likely matches its current center of diversity, East Asia. We found that, to explain the distribution of extant species in Oleaceae, the best biogeographic model given these data is one that incorporates the combined effects of the uplift of the QTP, which likely started in the Eocene and continued with bursts of rapid uplift in the Miocene, and those of the climatic variations in the Pliocene and Pleistocene (QTP-Climate hypothesis). This model also explains the East-West Eurasian disjunctions seen in *Fontanesia*, *Forsythia*, *Ligustrum*, *Osmanthus*, and *Syringa*, where it estimates that extinction, vicariance, and founder events mostly happened from the Early Miocene onwards.

The findings we present here are relevant steps to better understand the past diversification of Oleaceae. This phylogenetic framework, and previous ones based on earlier versions of this dataset (Dupin *et al*., 2020), has already helped to advance our knowledge on the origin and evolution of specific traits in Oleaceae such as the genetic determinants of self-incompatibility systems in the family (Castric *et al*., 2024; Raimondeau *et al*., 2024). We expect to see further efforts not only to disentangle the early biogeographic history of the family based on these findings, but also to improve knowledge of infrafamilial relationships, especially within under-sampled and/or large genera (e.g., *Chionanthus*, *Jasminum*, *Tetrapilus*) and within non-monophyletic taxa such as *Chionanthus* and *Phillyrea-Picconia*.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following:

Figure S1. Best ML tree of Oleaceae with all accessions represented.

Figure S2. Best ML tree of Oleaceae with all accessions and node support information.

Figure S3. Full results of ancestral range estimation in Oleaceae.

Figure S4. Evolution of the number of ranges that comprised at least one Asian area and Europe.

Table S1. List of accessions, along with biogeographic distribution.

Table S2. Best partition scheme, partition details, and best-fit model per partition according to Bayesian information criterion (BIC) implemented in ModelFinder

Table S3. Comparison of all 24 biogeographic models (four hypotheses, six models per hypothesis) using AIC and Akaike weights (Burnham and Anderson, 2002) to rank the fit of each model to the data.

Table S4. Mean counts, and their standard deviation values (in parentheses), of all dispersal events between areas, summarized from 100 BSMs. Rows represent areas from where dispersal started, and columns where it ended.

DATA AVAILABILITY

The data underlying this article (tree files, and input files and code for biogeographic analyses) are available in Figshare, at https://figshare.com/s/b91d6baf9ea72a5e0d5a. Sequences newly generated for this study are available in the GenBank Nucleotide Database, and can be accessed with the unique identifiers listed in the Supplementary Table S1.

FUNDING

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie actions (project FRUITFUL, grant agreement H2020-MSCA-IF-2018-842234). In addition, J.D. and G.B. are members of the EDB laboratory, which is supported by the excellence projects LabEx CEBA (ANR-10-LABX-25-01) and LabEx TULIP (ANR-10- LABX-0041), managed by the French "Agence Nationale de la Recherche".

ACKNOWLEDGEMENTS

We thank the herbaria staff at BONN, G, K, M, MO, MPU, P, PET, QBG, TAN, TUM, US, Laure Civeyrel, Olivier Maurin, Thierry Otto, and Terry D. Pennington for providing herbarium samples included in this work. We are also grateful to the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi:10.15454/1.5572369328961167E12) for providing computing and storage resources. We are grateful to Sophie Manzi for lab assistance (library construction).

Author contributions: J.D. and G.B. designed the study, established the sampling, did the lab work, and the data analyses. J.D. led the writing of the manuscript with significant contributions from C.H.-W., M.G., and G.B.

Declaration of interests: All authors have seen and agree with the contents of the manuscript.

LITERATURE CITED

Allen JM, LaFrance R, Folk RA, Johnson KP, Guralnick RP. 2018. aTRAM 2.0: an improved, flexible locus assembler for NGS data. *Evolutionary Bioinformatics* **14**: 1176934318774546.

Baas P, Esser PM, van der Westen MET, Zandee M. 1988. Wood anatomy of the Oleaceae. *IAWA Bulletin* **9**: 103–182.

Barrón E. 1992. Presencia de *Fraxinus excelsior* Linne (Oleaceae, Gentianales) en el Mioceno Superior de la depresión Ceretana: implicaciones tafonóicas y paleoecológicas. *Revista Española de Paleontología* **7**: 101–108.

Battandier JB, Trabut L. 1911. Contribution à la Flore du pays des Touaregs. *Bulletin de la Société Botanique de France* **58**: 669–677.

Besnard G, Rubio de Casas R, Christin P-A, Vargas P. 2009. Phylogenetics of *Olea* (Oleaceae) based on plastid and nuclear ribosomal DNA sequences: Tertiary climatic shifts and lineage differentiation times. *Annals of Botany* **104**: 143–160.

Bezanson J, Edelman A, Karpinski S, Shah VB. 2017. Julia: a fresh approach to numerical computing. *SIAM Review* **59**: 65–98.

Bouckaert R, Vaughan TG, Barido-Sottani J, *et al.* 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **15**: e1006650.

Browicz K. 1992. East-West Euroasiatic disjunction of woody genera. *Arboretum Kórnickie* **37**: 5–19.

Burnham KP, Anderson DR. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* 2nd Edition. New York: Springer-Verlag.

Cai R, Ané C. 2021. Assessing the fit of the multi-species network coalescent to multi-locus data. *Bioinformatics* **37**: 634–641.

Call VB, Dilcher DL. 1992. Investigations of angiosperms from the Eocene of southeastern North America: samaras of *Fraxinus wilcoxiana* Berry. *Review of Palaeobotany and Palynology* **74**: 249–266.

Castric V, Batista RA, Carré A, *et al.* 2024. The homomorphic self-incompatibility system in Oleaceae is controlled by a hemizygous genomic region expressing a gibberellin pathway gene. *Current Biology* **34**: 1967-1976..

Chen YS, Meseguer AS, Godefroid M, *et al.* 2017. Out-of-India dispersal of *Paliurus* (Rhamnaceae) indicated by combined molecular phylogenetic and fossil evidence. *Taxon* **66**: 78–90.

Chernomor O, Von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology* **65**: 997–1008.

Cohen KM, Harper DAT, Gibbard PL. 2022. ICS International Chronostratigraphic Chart YYYY/MM. International Commission on Stratigraphy, IUGS. www.stratigraphy.org (visited: 2021/10/01).

Dierckxsens N, Mardulyn P, Smits G. 2017. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* **45**: e18.

Djamali M, Brewer S, Breckle S-W, Jackson ST. 2012. Climatic determinism in phytogeographic regionalization: A test from the Irano-Turanian region, SW and Central Asia. *Flora* **207**: 237–249.

Dong W, Li E, Liu Y, *et al.* 2022. Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family. *BMC Biology* **20**: 92.

Dupin J, Hong-Wa C, Pillon Y, Besnard G. 2022. From the Mediterranean to the Pacific: re-circumscription towards *Notelaea s.l.* and historical biogeography of a generic complex within Oleinae (Oleaceae). *Botanical Journal of the Linnean Society* **200**: 360–377.

Dupin J, Matzke NJ, Särkinen T, *et al.* 2017. Bayesian estimation of the global biogeographical history of the Solanaceae. *Journal of Biogeography* **44**: 887–899.

Dupin J, Raimondeau P, Hong-Wa C, Manzi S, Gaudeul M, Besnard G. 2020. Resolving the phylogeny of the olive family (Oleaceae): confronting information from organellar and nuclear genomes. *Genes* **11**: 1508.

Fiz-Palacios O, Vargas P, Vila R, Papadopulos AST, Aldasoro JJ. 2010. The uneven phylogeny and biogeography of *Erodium* (Geraniaceae): radiations in the Mediterranean and recent recurrent intercontinental colonization. *Annals of Botany* **106**: 871–884.

Goodstein DM, Shu S, Howson R, *et al.* 2018. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**: D1178–D1186.

Green AF, Ramsey TS, Ramsey J. 2011. Phylogeny and biogeography of ivies (*Hedera* spp., Araliaceae), a polyploid complex of woody vines. *Systematic Botany* **36**: 1114–1127.

Green PS. 1972. *Osmanthus decorus* and disjunct Asiatic-European distributions in the Oleaceae. *Kew Bulletin* **26**: 487–490.

Green PS. 2001. Studies in the genus *Jasminum*, XVII: sections *Trifoliata* and *Primulina*. *Kew Bulletin* **56**: 903–915.

Green PS. 2002. A revision of *Olea* L. (*Oleaceae*). *Kew Bulletin* **57**: 91–140.

Green PS. 2004. Oleaceae. In: Kubitzki K, Kadereit JW, eds. *The Families and Genera of Vascular Plants*. Vol. VII: *Flowering Plants, Dicotyledons*. New York: Springer, 296–306.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology* **59**: 307–321.

Guo ZT, Sun B, Zhang ZS, *et al.* 2008. A major reorganization of Asian climate by the early Miocene. *Climate of the Past* **4**: 153–174.

Ha YH, Kim C, Choi K, Kim JH. 2018. Molecular phylogeny and dating of Forsythieae (Oleaceae) provide insight into the Miocene history of Eurasian temperate shrubs. *Frontiers in Plant Science* **9**: 99.

Harborne JB, Green PS. 1980. A chemotaxonomic survey of flavonoids in leaves of the Oleaceae. *Botanical Journal of the Linnean Society* **81**: 155–167.

Harrison TM, Copeland P, Kidd WSF, Yin AN. 1992. Raising Tibet. *Science* **255**: 1663–1670.

Hinsinger DD, Basak J, Gaudeul M, *et al.* 2013. The phylogeny and biogeographic history of ashes (*Fraxinus*, Oleaceae) highlight the roles of migration and vicariance in the diversification of temperate trees. *PLoS One* **8**: e80431.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518–522.

Hong-Wa C, Besnard G. 2013. Intricate patterns of phylogenetic relationships in the olive family as inferred from multi-locus plastid and nuclear DNA sequence analyses: A close-up on *Chionanthus* and *Noronhia* (Oleaceae). *Molecular Phylogenetics and Evolution* **67**: 367–378.

Hong-Wa C, Dupin J, Frasier C, Schatz GE, Besnard G. 2023. Systematics and biogeography of Oleaceae subtribe Schreberinae, with recircumscription and revision of the Malagasy members. *Botanical Journal of the Linnean Society* **202**: 476–509.

John KES, Krissek LA. 2002. The Late Miocene to Pleistocene ice-rafting history of southeast Greenland. *Boreas* **31**: 28–35.

Johnson LAS. 1957. A review of the family Oleaceae. *Contributions from the New South Wales National Herbarium* **2**: 395–418.

Johnson MG, Pokorny L, Dodsworth S, *et al.* 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* **68**: 594–606.

de Juana Clavero JI. 2020. Propuesta de *Tetrapilus* Lour. (Oleaceae) como género válido. Especies presentes en España. *Bouteloua* **30**: 54–60.

Landis MJ, Matzke NJ, Moore BR, Huelsenbeck JP. 2013. Bayesian analysis of biogeography when the number of areas is large. *Systematic Biology* **62**: 789–804.

Landis MJ. 2017. Biogeographic dating of speciation times using paleogeographically informed processes. *Systematic Biology* **66**: 128–144.

Lee H-L, Jansen RK, Chumley TW, Kim K-J. 2007. Gene relocations within chloroplast genomes of *Jasminum* and *Menodora* (Oleaceae) are due to multiple, overlapping inversions. *Molecular Biology and Evolution* **24**: 1161–1180.

Li J, Alexander JH, Zhang D. 2002. Paraphyletic *Syringa* (Oleaceae): evidence from sequences of nuclear ribosomal DNA ITS and ETS regions. *Systematic Botany* **27**: 592–597.

Li J, Goldman-Huertas B, DeYoung J, Alexander J. 2012. Phylogenetics and diversification of *Syringa* inferred from nuclear and plastid DNA sequences. *Castanea* **77**: 82–88.

Li Y-F, Zhang M, Wang X-R, *et al.* 2020. Revisiting the phylogeny and taxonomy of *Osmanthus* (Oleaceae) including description of the new genus *Chengiodendron*. *Phytotaxa* **436**: 283–292.

Lu Y, Dewald N, Koutsodendris A, *et al.* 2020. Sedimentological evidence for pronounced glacial-interglacial climate fluctuations in NE Tibet in the latest Pliocene to Early Pleistocene. *Paleoceanography and Paleoclimatology* **35**: e2020PA003864.

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

Manafzadeh S, Salvo G, Conti E. 2014. A tale of migrations from east to west: the Irano-Turanian floristic region as a source of Mediterranean xerophytes. *Journal of Biogeography* **41**: 366–379.

Manafzadeh S, Staedler YM, Conti E. 2017. Visions of the past and dreams of the future in the Orient: the Irano-Turanian region from classical botany to evolutionary studies. *Biological Reviews* **92**: 1365–1388.

Martínez-Millán M. 2010. Fossil record and age of the Asteridae. *The Botanical Review* **76**: 83–135.

Mathewes R, Archibald SB, Lundgren A. 2021. Tips and identification of Early Eocene *Fraxinus* L. samaras from the Quilchena locality, Okanagan Highlands, British Columbia, Canada. *Review of Palaeobotany and Palynology* **293**: 104480.

Matzke NJ. 2014. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* **63**: 951–970.

Matzke NJ. 2018a. *BioGeoBEARS: BioGeography with Bayesian (and likelihood) Evolutionary Analysis with R Scripts*. version 1.1.1. published on GitHub on November 6, 2018.

Matzke NJ. 2018b. *cladoRcpp v0.15.1: C++ Implementations of Phylogenetic Cladogenesis Calculations.* University of Auckland, New Zealand.

Matzke N, Sidje R, Schmidt D. 2019. *rexpokit v0.26.6.6: R wrappers for EXPOKIT; other matrix functions.* School of Biological Sciences, University of Auckland, New Zealand.

Minh BQ, Schmidt HA, Chernomor O, *et al.* 2020a. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Minh BQ, Hahn MW, Lanfear R. 2020b. New methods to calculate concordance factors for phylogenomic datasets. *Molecular Biology and Evolution* **37**: 2727–2733.

Morrone JJ. 2015. Biogeographical regionalisation of the world: a reappraisal. *Australian Systematic Botany* **28**: 81–90.

Olofsson JK, Cantera I, Van de Paer C, *et al.* 2019. Phylogenomics using low-depth whole genome sequencing: A case study with the olive tribe. *Molecular Ecology Resources* **19**: 877–892.

Pei NC, Li Y. 2024. The golden bell flower genome provides insights into its evolutionary history and reveals the potential genomic bases of its ecological divergence from weeping forsythia. *Scientia Horticulturae* **323**: 112527.

Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution* **21**: 1455–1458.

POWO. 2023. Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://www.plantsoftheworldonline.org/. Retrieved 24 September 2023.

Qiao C, Ran J, Li Y, Wang X. 2007. Phylogeny and biogeography of *Cedrus* (Pinaceae) inferred from sequences of seven paternal chloroplast and maternal mitochondrial DNA regions. *Annals of Botany* **100**: 573–580.

Raimondeau P, Ksouda S, Marande W, *et al.* A hemizygous supergene controls homomorphic and heteromorphic self-incompatibility systems in the olive family (Oleaceae). *Current Biology* **34**: 1977-1986.

Ree RH, Smith S. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* **57**: 4–14.

Renne PR, Deino AL, Hilgen FJ, *et al.* 2013. Time scales of critical events around the Cretaceous-Paleogene boundary. *Science* **339**: 684–7.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology* **46**: 195–203.

Rohwer JG. 1997. The fruits of *Jasminum mesnyi* (Oleaceae), and the distinction between *Jasminum* and *Menodora*. *Annals of the Missouri Botanical Garden* **84**: 848–856.

Sauquet H. 2013. A practical guide to molecular dating. *Comptes Rendus Palevol* **12**: 355–367.

Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689–2690.

Solís-Lemus C, Ané C. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genetics* **12**: e1005896.

Solís-Lemus C, Bastide P, Ané C. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* **34**: 3292–3298.

Stearn WT. 1976. Union of *Chionanthus* and *Linociera* (Oleaceae). *Annals of the Missouri Botanical Garden* **63**: 355–357.

Stearn WT. 1980. African species of *Chionanthus* L. (Oleaceae) hitherto included in *Linociera* Swartz. *Botanical Journal of the Linnean Society* **80**: 191–206.

Steyermark JA. 1932. A revision of the genus *Menodora*. *Annals of the Missouri Botanical Garden* **19**: 87–176.

Sun H. 2002. Tethys retreat and Himalayas-Hengduanshan Mountains uplift and their significance on the origin and development of the Sino-Himalayan elements and alpine flora. *Acta Botanica Yunnanica* **24**: 273–288.

Sun H, McLewin W, Fay MF. 2001. Molecular phylogeny of *Helleborus* (Ranunculaceae), with an emphasis on the eastern Asian-Mediterranean disjunction. *Taxon* **50**: 1001–1018.

Tiffney BH, Manchester SR. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. *International Journal of Plant Sciences* **162**: S3– S17.

Tillich M, Lehwark P, Pellizzer T, *et al.* 2017. GeSeq: versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* **45**: W6–W11.

Tu T, Volis S, Dillon MO, Sun H, Wen J. 2010. Dispersals of Hyoscyameae and Mandragoreae (Solanaceae) from the New World to Eurasia in the Early Miocene and their biogeographic diversification within Eurasia. *Molecular Phylogenetics and Evolution* **57**: 1226–1237.

Wallander E, Albert VA. 2000. Phylogeny and classification of Oleaceae based on *rps16* and *trnL-F* sequence data. *American Journal of Botany* **87**: 1827–1841.

Webb III T, Bartlein PJ. 1992. Global changes during the last 3 million years: climatic controls and biotic responses. *Annual review of Ecology and Systematics* **23**: 141–173.

Wu MX, Huang J, Su T, Zhou Z K, Xing, Y W. 2021. *Fraxinus* L. (Oleaceae) fruits from the Early Oligocene of Southwest China and their biogeographic implications. *Fossil Imprint* **77**: 287–298.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**: 686–693.

Zhang Z, Fan L, Yang J, Hao X, Gu Z. 2006. Alkaloid polymorphism and ITS sequence variation on the *Spiraea japonica* complex (Rosaceae) in China: traces of the biological effects of the Himalaya-Tibet Plateau uplift. *American Journal of Botany* **93**: 762–769.

TABLE 1. Number and proportion of species of Oleaceae found today in the main biogeographic areas of this study (full species list) and sampled for our biogeographic analysis (biogeographic tree). Numbers correspond to species whose native distribution match fully or partially the region.

	Full species list (736 spp.)		**Biogeographic tree[a] (211 spp.)**	
Biogeographic areas	**n**	**%**	**n**	**%**
Temperate Asia	63	9	26	12
Tropical Asia	323	44	89	42
Australasia[b]	85	12	25	12
Europe	22	3	16	8
Africa	170	23	48	23
Northern America	62	8	17	8
Southern America	81	11	22	10

[a] Biogeographic tree: pruned version of our best ML tree, to approximate the proportions of the number of species per area to the full species list in Oleaceae; [b] Australasia includes the Pacific region and the Hawaiian Islands.

TABLE 2. Node support values and confidence factors for Oleaceae's tribes, subtribes, and clades between them.

Clade	UFBootstrap	SH-aLRT	gCF	sCF
Tribes of Oleaceae				
Myxopyreae	100	100	92.9	76.3
Fontanesieae	100	100	100	100
Forsythieae	100	100	100	88.4
Jasmineae	100	100	100	94.5
Oleeae	100	100	85.7	71.2
(Jasmineae, Oleeae)	100	100	57.1	48.0
(Forsythieae, (Jasmineae, Oleeae))	93	19.4	21.4	27.1
(Fontanesieae, (Forsythieae, (Jasmineae, Oleeae)))	100	100	64.3	61.6
(Myxopyreae, (Fontanesieae, (Forsythieae, (Jasmineae, Oleeae))))	100	100	80.0	84.5
Subtribes of Oleeae				
Oleinae	100	100	64.3	64.6
Fraxininae	100	100	70.0	70.9
Ligustrinae	100	100	60.0	72.1
Schreberinae	100	100	80.0	71.8
(Fraxininae, Oleinae)	100	100	57.1	57.8
(Ligustrinae, (Fraxininae, Oleinae))	100	100	50.0	21.6

TABLE 3. Crown node ages, with the upper and lower bounds of the 95% higher posterior density (HPD), for Oleaceae and selected groups within it, under the ORC molecular clock.

Clade	median	95% HPD min	95% HPD max
Oleaceae	85.9	83.7	87.2
(Fontanesieae, (Forsythieae, (Jasmineae, Oleeae)))	80.0	76.5	83.3
(Forsythieae, (Jasmineae, Oleeae))	78.0	74.3	81.4
(Jasmineae, Oleeae)	73.9	69.9	77.5
Myxopyreae	68.0	60.2	75.3
Fontanesieae	0.3	0	1.5
Forsythieae	13.6	10.1	17.7
Jasmineae	52.3	46.6	57.4
Oleeae	56.7	54.1	59.9
(Ligustrinae, (Fraxininae, Oleinae))	53.0	50.7	55.7
(Fraxininae, Oleinae)	48.5	47.9	49.7
Schreberinae	35.1	26.1	44.8
Ligustrinae	32.5	23.6	42.9
Fraxininae	41.0	37.1	45.5
Oleinae	33.7	30.1	37.8

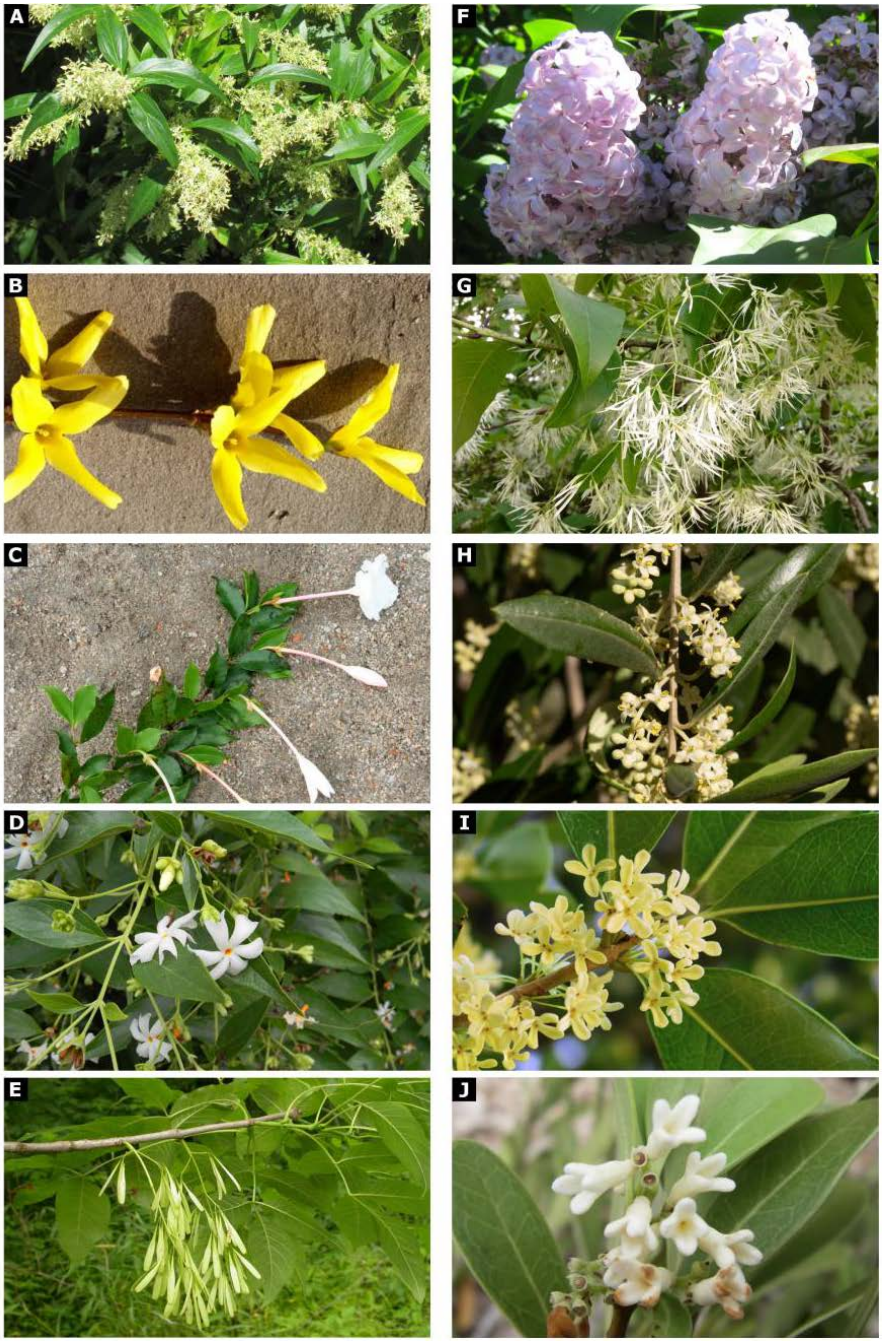

Figure 1. Representatives of tribes and subtribes in Oleaceae. Tribe Fontanesieae: (A) *Fontanesia fortune*; tribe Forsythieae: (B) *Forsythia intermedia*; tribe Jasmineae: (C) *Jasminum kitchingii* (©Anton Sieder, CC BY-NC-ND 3.0); tribe Myxopyreae: (D) *Nyctanthes arbor-tristis* (©Varun Pabrai, CC BY-SA 4.0); tribe Oleeae, subtribe Fraxininae: (E) *Fraxinus pennsylvanica* (©Gerrit Davidse, CC BY-NC-SA 3.0); subtribe Ligustrinae: (F) *Syringa vulgaris*; subtribe Oleinae: (G) *Chionanthus virginicus*, (H) *Olea europaea* (©Nefronus, CC BY-SA 4.0), (I) *Osmanthus fragrans* (©Junichiro Aoyama, CC BY 2.0); subtribe Schreberinae: (J) *Schrebera minor*. Images without copyright licenses listed were provided by authors.

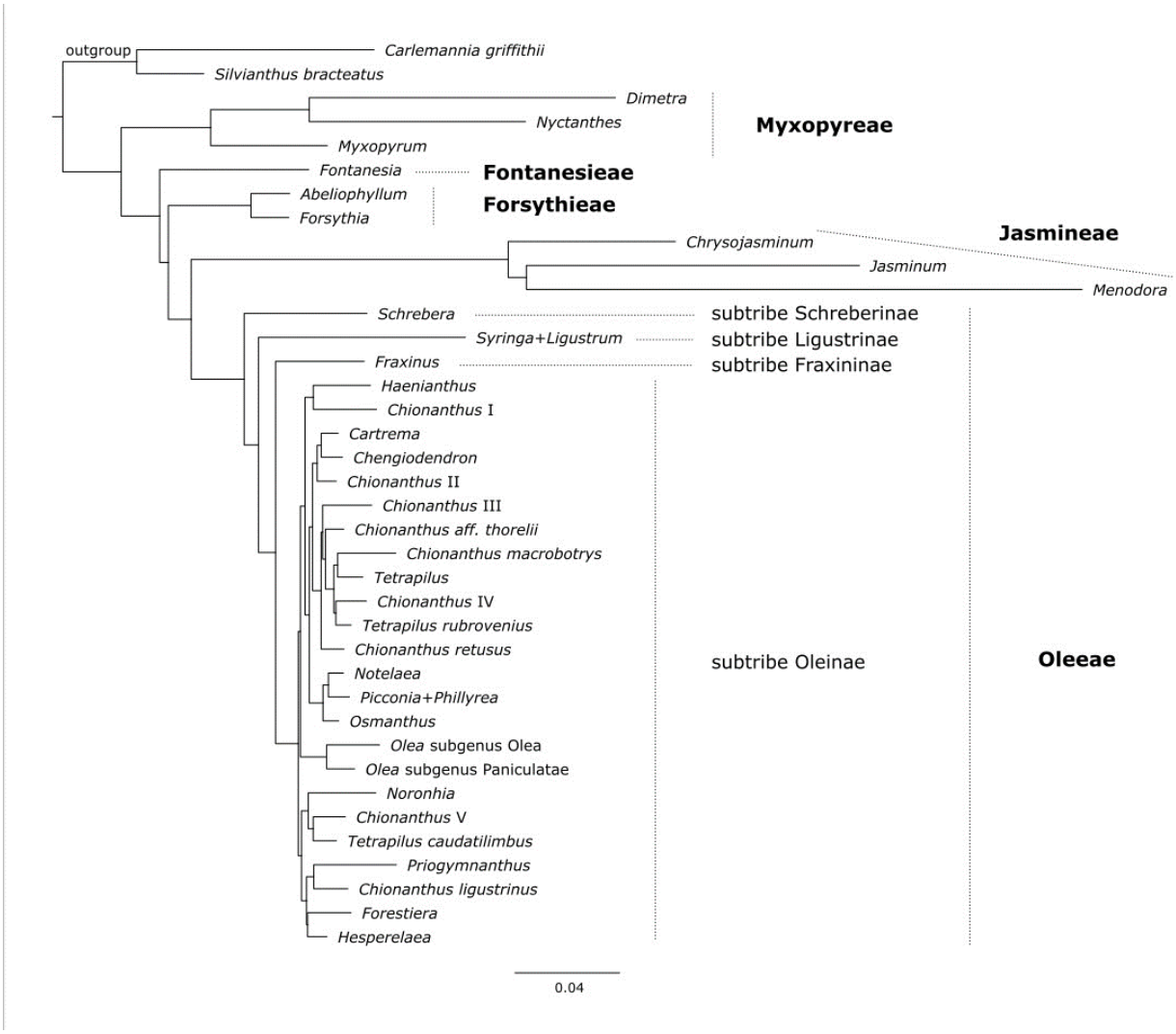

Figure 2. Simplified topology of the best ML tree of Oleaceae and outgroups. This tree results from an analysis of a concatenated dataset of plastid and nuclear regions. Tribe names are in bold. For the full tree, where all accessions are shown, please refer to Supplementary Figure S1. For a tree with information on node support, please see Supplementary Figure S2.

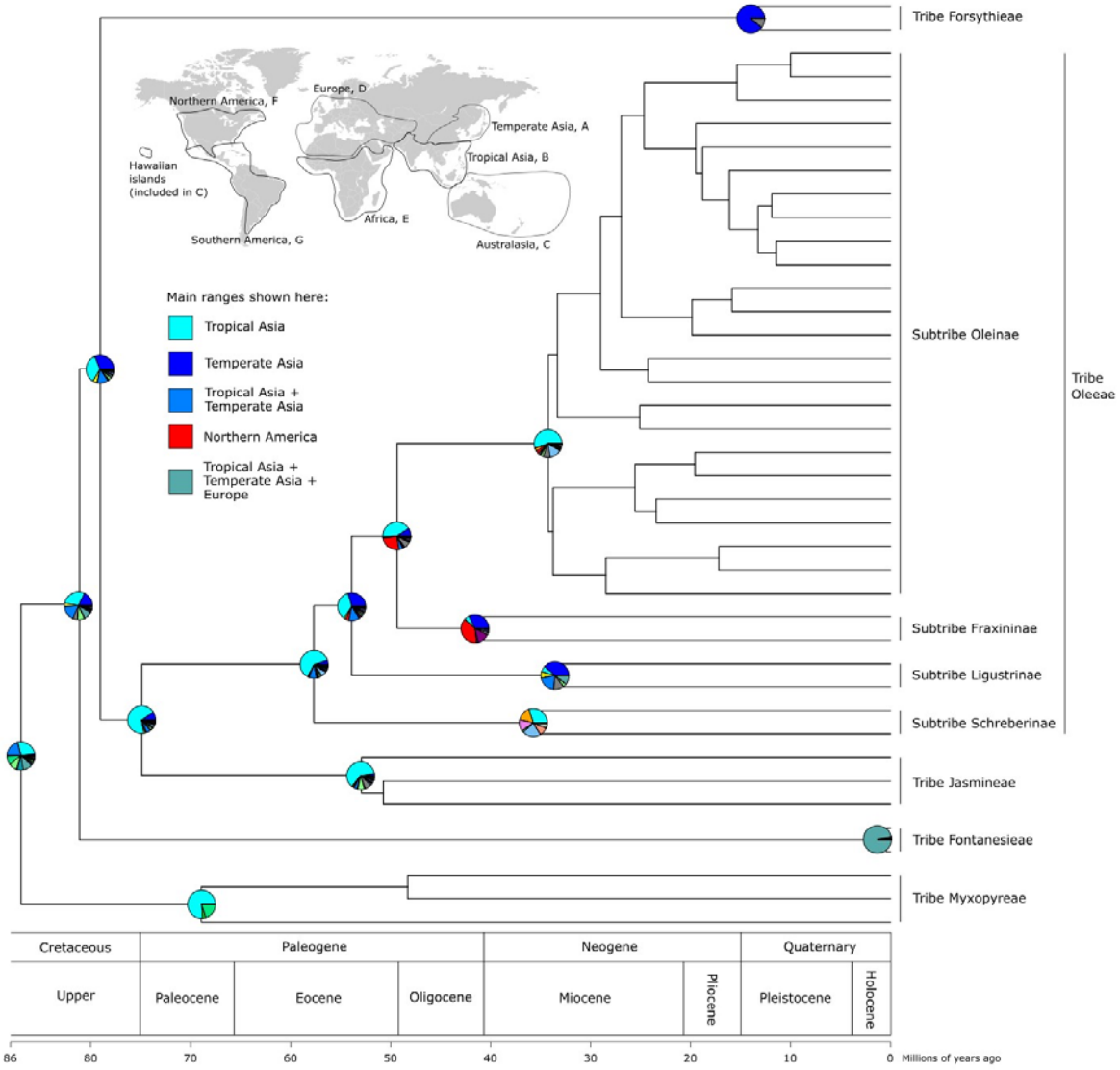

Figure 3. Maximum Likelihood ancestral range estimation in Oleaceae, using the best model DEC+*j* (model 14 in Supplementary Table S3). This tree is a simplified version of our 402-tip tree (Supplementary Figure S3), and it represents tribes and subtribes within Oleaceae. The pie diagrams at nodes show the relative probability of the possible states (ranges). The boxes on the left show the most common ranges depicted in this tree.

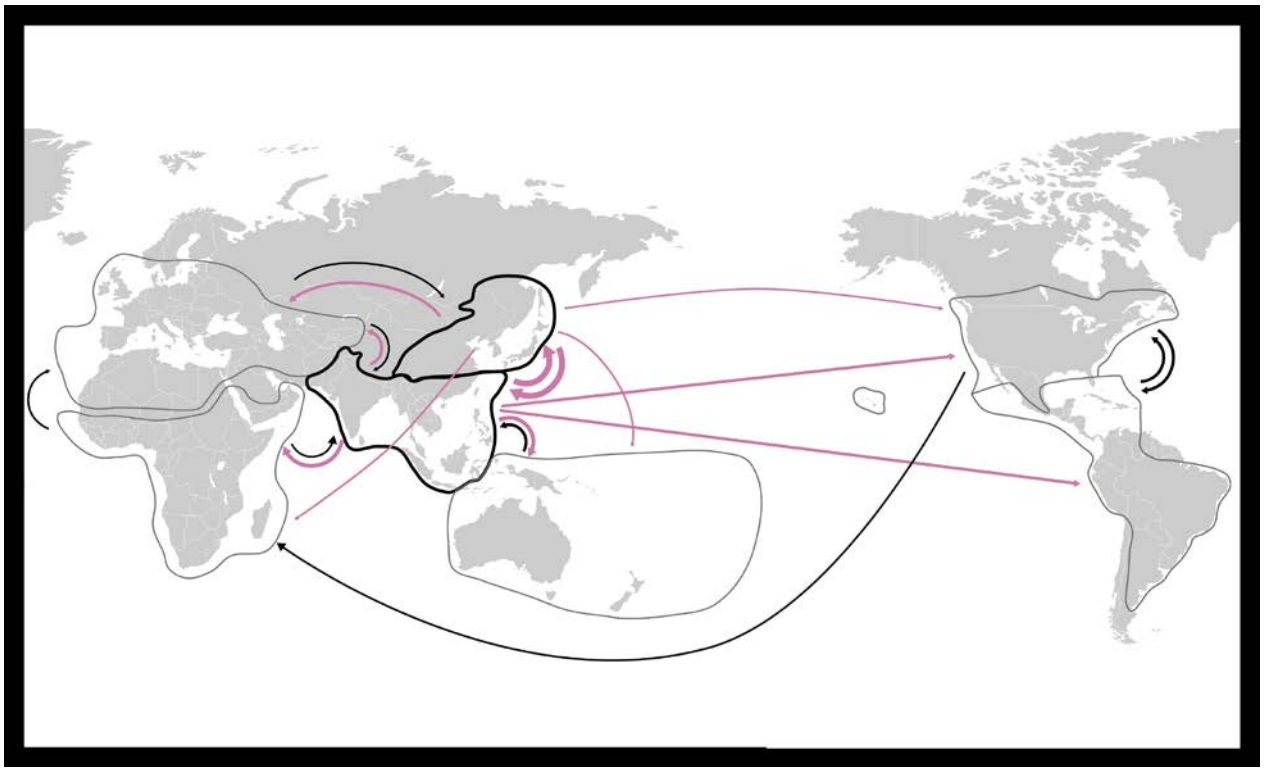

Figure 4. Summary of dispersal events estimated with biogeographical stochastic mapping in the history of Oleaceae. Contoured areas are the same ones as seen in Fig. 3. The two areas with a bold black contour, Tropical and Temperate Asia, are the most species-rich today, and events from those areas are highlighted in pink. The arrows between areas represent direction and frequency of dispersal events. Only event counts that presented a mean of 0.95 or higher (Supplementary Table S4) are depicted as arrows here; arrow line thickness is proportional to the mean number of events.

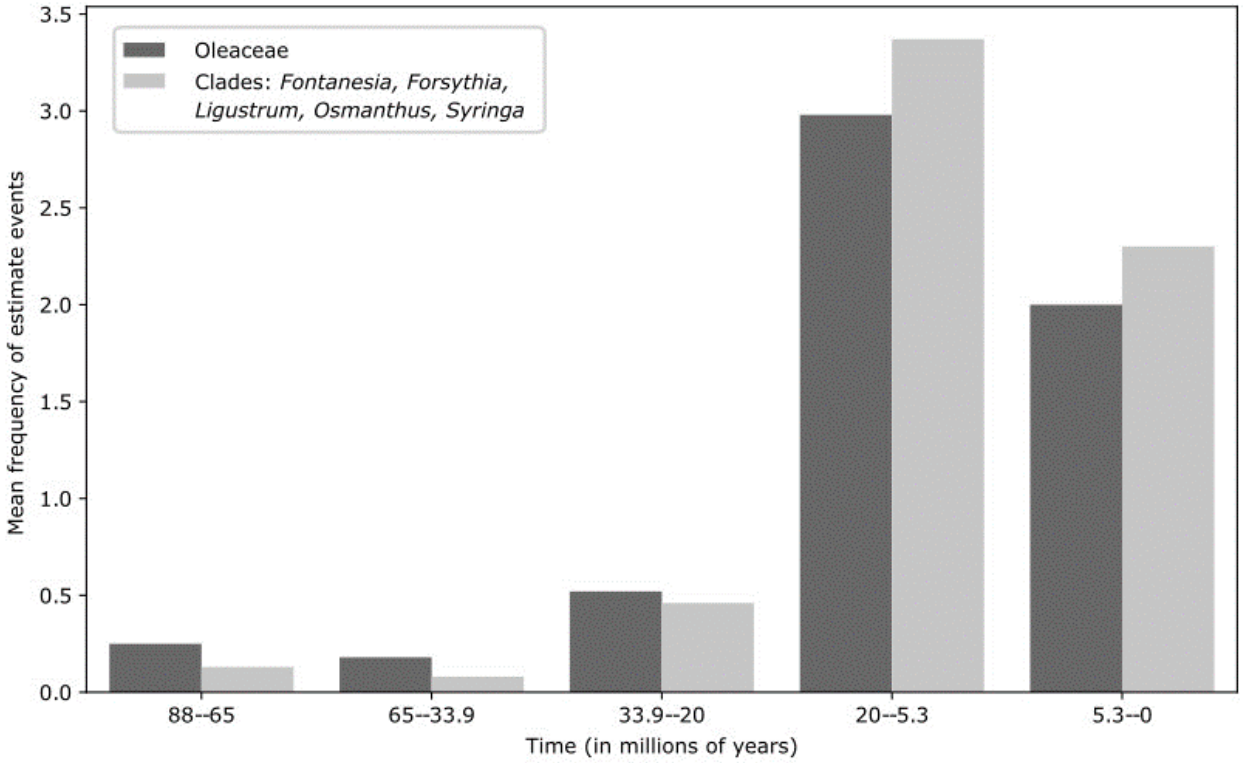

Figure 5. Timing and frequency of estimated extinction, vicariance, and founder events that likely contributed to disjunctions between Asia and Europe. For the sake of comparison, we show the frequency of such events in the entire family, and the frequency only in the clades with East-West Eurasian disjunctions. Time is represented in five categories: from 88 to 65 Mya, 65 to 33.9 Mya, 33.9 to 20 Mya, 20 to 5.3 Mya, and 5.3 Mya to present time.