# High quality long-read genomes produced from single MinION flow cells clarify polyploid and demographic histories of critically endangered ash species (Fraxinus: Oleaceae)

**Steven Fleck**
**Crystal Tomlin**
**Flavio da Silva Coelho**
**Michaela Richter**
**Erik Danielsen**
**Nathan Backenstose**
https://orcid.org/0000-0002-4793-704X
**Trevor Krabbenhoft**
University at Buffalo https://orcid.org/0000-0002-7680-5169
**Charlotte Lindqvist**
University at Buffalo (SUNY) https://orcid.org/0000-0002-4190-727X
**Victor Albert** ( **vaalbert@buffalo.edu** )
University at Buffalo https://orcid.org/0000-0002-0262-826X

**Article**

**Keywords:**

**Posted Date:** December 14th, 2022

**DOI:** https://doi.org/10.21203/rs.3.rs-2350866/v1



**Additional Declarations:** There is **NO** Competing Interest.

**Version of Record:** A version of this preprint was published at Communications Biology on January 6th, 2024. See the published version at https://doi.org/10.1038/s42003-023-05748-4.

# Abstract

With populations of threatened and endangered plants and animals declining worldwide, it is important that high quality genomic records of these species are preserved before they are lost forever. Here, we demonstrate that data from single Oxford Nanopore Technologies (ONT) MinION flow cells can, even in the absence of highly accurate short DNA-read polishing, produce high quality *de novo*plant genome assemblies that are adequate for downstream analyses, such as synteny and ploidy evaluations, paleodemographic analyses, and phylogenomics. This study focuses on three North American ash tree species in the genus *Fraxinus*(Oleaceae) that were recently added to the International Union for Conservation of Nature (IUCN) Red List: *Fraxinus americana* (white ash), *F. nigra* (black ash), and *F. pennsylvanica* (green ash). These three species have become critically endangered primarily due to destructive herbivory by the invasive Emerald Ash Borer (EAB, *Agrilus planipennis*), a buprestid beetle pest recently introduced to North America from East Asia. Our draft genomes, which range from 776.3-851.9 megabases, have similar sequence accuracy as a recently published chromosome-level *F. pennsylvanica* assembly, with annotations that outperform that genome in terms of the number of complete Benchmarking Universal Single-Copy Orthologs (BUSCOs) identified. Our results support a whole genome triplication at the base of the Oleaceae as well as a subsequent whole genome duplication shared by *Syringa*, *Osmanthus*, *Olea, and Fraxinus*. Additionally, our results from ONT long reads alone suggest that our *F. nigra* accession is more inbred compared with the *F. americana* and *F. pennsylvanica* individuals sequenced. In summary, our powerful downstream analyses enabled by single MinION flow cell genome assemblies suggest that Oxford Nanopore technology can provide a relatively fast and inexpensive approach to sequence the 5,232 critically endangered plant species currently on the IUCN Red-List.


# Significance

We present three high quality critically endangered ash (*Fraxinus*) genome assemblies generated solely from long-read sequence data from single Oxford Nanopore MinION flow cells. We corroborate a whole genome triplication (WGT) at the base of the Oleceae present in *Jasminum* and *Forsythia* and the subsequent whole genome duplication shared by lilac (*Syringa*), sweet osmanthus (*Osmanthus*), olive (*Olea*), and ash (*Fraxinus*). We also show that ONT long reads can be used to detect differences in paleodemography through time using Pairwise Sequentially Markovian Coalescent (PSMC) analyses.

# Introduction

With populations of threatened and endangered organisms continually declining, it is vital that high quality genomic records of these species are preserved before they are lost forever. In 2017, five North American ash trees (*Fraxinus* L.; Oleaceae) were added to the International Union for Conservation of Nature (IUCN) Red List as critically endangered: white ash (*Fraxinus americana* L.) [1], black ash (*F. nigra* Marsh.) [2], green ash (*F. pennsylvanica* Marsh.) [3], pumpkin ash (*F. profunda* Bush) [4], and blue ash (*F. quadrangulata* Michx.) [5]. These North American ash trees have become critically endangered due to the introduction of the emerald ash borer (EAB, *Agrilus planipennis* Fairmaire), a buprestid beetle [6]. The EAB

was first identified in southeast Michigan in 2002 and has since spread to most of the United States, resulting in the death of millions of ash trees as well as massive economic and ecological impacts [6]. The larval stage of the EAB feeds on the phloem of ash trees, girdling them as adults, or even long before the tree reaches maturity (as small as 2.5 cm diameter at breast height; dbh), which kills the tree within 2–4 years of infestation [6]. North American ash species have shown some low level of resistance [7], but *F. americana*, *F. nigra*, and *F. pennsylvanica* are highly vulnerable to the EAB, with some studies showing virtually 100% mortality and newly germinated ash seedlings scarce or completely absent [8].

Recent progress has been made on the evolution of EAB resistance in *Fraxinus*. Kelly et al. [7] identified 53 candidate *Fraxinus* genes for EAB resistance and demonstrated that 48 likely arose through convergent evolution in three separate lineages (taxonomic sections) within the genus: *F. mandshurica* Rupr., a member of the *Fraxinus* section; *F. platypoda* Oliv., sister to the *Melioides* section; *F. baroniana* Diels, *F. floribunda* Wall., and *Fraxinus sp.* D2006-0159, clustered within section *Ornus*. All five EAB-resistant species are native to Asia and within the native range of the EAB, except for the present-day ranges of *F. baroniana* and *F. floribunda*. Loss-of-function mutations for candidate genes were more common in ash trees that were susceptible to the EAB than species that were resistant. For example, *F. pennsylvanica*, another member of section *Melioides*, displays loss of function in one candidate gene due to two frame shift mutations and a loss of function mutation in another candidate gene due to a stop codon gain [7]. *F. pennsylvanica* also includes rare, putatively EAB-resistant individuals which lack these loss of function mutations. Out of the fifty-three candidate genes, twenty-four have putative roles relating to defense response against insect herbivory, including phytohormone biosynthesis and signaling, and others have potential roles related to defense response.

Recent progress has also been made in comparative genomics of the Oleaceae. High quality, chromosome-level genome assemblies have been published for Arabian jasmine (*Jasminum sambac* (L.) Aiton) [9], sweet osmanthus (*Osmanthus fragrans* Lour. var. 'Rixianggui' [10] and the cultivar *O. fragrans* 'Liuyejingui' [11]), weeping forsythia (*Forsythia suspensa* (Thunb.) Vahl) [12, 13], common olive (*Olea europaea* L.) [14], and early lilac (*Syringa oblata* Lindl.) [15, 16]. For the genus *Fraxinus*, Kelly et al. [7] published twenty-eight short-read *de novo* assemblies, representing twenty-two species-level taxa, including three *F. pennsylvanica* assemblies, one EAB-susceptible (pe-48), one partially EAB-resistant (pe-00248), and one originally misidentified as *F. caroliniana*. Additionally, three *F. angustifolia* Vahl subspecies and two unidentified *Fraxinus* species were sequenced [7]. Huff et al. [17] generated new 800 basepair (bp) insert size library data for eight of these assemblies, as well as for an additional species, *F. apertisquamifera* H.Hara, and created new short-read *de novo* assemblies. They also generated a chromosome-level assembly for the putatively EAB-resistant *F. pennsylvanica* individual. These twenty-six short-read assemblies, along with the previously published *F. excelsior* L. [18], were scaffolded using the *F. pennsylvanica* reference assembly as a guide [19]. The resulting pseudo-chromosome-level assemblies were able to place about 45%-87% of the total lengths of the short-read input assemblies into pseudomolecules.

Polyploidy events have been considered important for the evolution of the Oleaceae. Within the genome of *Jasminum sambac*, roughly 30% of genes in rapidly expanding gene families were associated with a whole genome triplication (WGT) event [9]. These adaptive hypotheses involve duplicates of numerous stress-related and secondary metabolism associated genes, with the evolution of floral fragrance thought to be influenced by polyploidization and tandem duplication. It is possible for polyploidy events to increase diversification rates and species richness [20], and the proposed WGT event at the base of the Oleaceae may have contributed to the diversification and adaptive potential of the Oleaceae since its divergence from other Lamiales [9]. Likewise, a more recent whole genome duplication (WGD) event shared by all members of the Oleeae tribe may have been important for the evolution of these species. In *Osmanthus fragrans* 'Rixianggui', WGD and tandem gene duplication were hypothesized to have a role in the evolution the monoterpene synthesis pathway that generates the strong, sweet aroma of its flowers [10]. Additionally, the functional divergence of oil biosynthesis pathway genes following duplication may have been important in the evolution of *Olea europaea var. sylvestris*, an important oil crop [21].

Here, we generate highly accurate and coding sequence-complete haploid genome assemblies for the critically endangered species white ash (*Fraxinus americana*), black ash (*F. nigra*), and green ash (*F. pennsylvanica*) as part of the Org.ONE project. Org.ONE is a pilot-stage project initiated by Oxford Nanopore Technologies (ONT) that aims to support faster and local ultra-long read DNA sequencing and subsequent de novo genome assemblies of critically endangered species (https://org.one/oo). Org.ONE provides free-of-charge consumables for sequencing the DNA of critically endangered species on the condition that the data is uploaded to the EBI public database (https://www.ebi.ac.uk/ena/browser/home). The three ash species were sequenced using single ONT MinION flow cells each. Advancements in homology-based gene prediction allowed us to produce high quality genome annotations without needing to produce RNA-seq reads for our samples. The resulting reads and genome assemblies were of sufficient accuracy to evaluate polyploidy level and paleodemographic history using long-read instead of the deep short-read sequence data typically employed.

## Results And Discussion

# Genome assembly and annotation

Totals of 6,065,809, 3,788,248 and 2,581,752 reads with N50s of 10,916, 11,213, and 34,165 were generated for wild-collected, Chautauqua County, New York *Fraxinus americana*, *F. nigra*, and *F. pennsylvanica* individuals, respectively (Table S1; Figure S1). Mean 1C value estimates for *F. americana*, *F. nigra*, and *F. pennsylvanica* were 0.895 pg (875 MB), 0.847 pg (829 MB), and 0.889 pg (869 MB), as calculated using data from Whittemore et al. [22]. Comparing estimated genome sizes with total bases generated, *F. americana*, *F. nigra*, and *F. pennsylvanica* had 25.0x, 22.4x, and 23.0x coverage, respectively.

The initial *de novo* genome assemblies for *Fraxinus americana*, *F. nigra*, and *F. pennsylvanica* had high levels of completeness based on Benchmarking Universal Single-Copy Orthologs (BUSCOs) [23], i.e.,

96.4%, 95.4%, and 96.0% complete BUSCOs from the embryophyta_odb10 database, respectively (Table S2). These completeness levels were very similar to the unannotated *F. pennsylvanica* reference assembly (v1.4) from Huff et al. [17] with 97.0% complete embryophyta_odb10 BUSCOs (Table S2). Repetitive elements made up nearly 60% of all three *Fraxinus* genome assemblies (Table S3), while they made up about only about 45% of the *F. pennsylvanica* reference assembly, which was based on Illumina short reads. Since short-read assemblies have the propensity to collapse duplicated sequences, this difference was likely due to better read-through of repetitive DNA on long ONT single molecules [24]. The largest group of retroelements discovered were Ty1/Copia and Gypsy/DIRS LTR elements, each comprising about 15% of all repetitive elements characterized.

Annotations were generated using GeMoMa [25, 26], a homology-based gene prediction software, using annotations from *Fraxinus excelsior* [18], *Olea europaea* [14], *Jasminum sambac* [9], and *Arabidopsis thaliana* [27]. The annotation of Huff et al.'s *F. pennsylvanica* reference assembly was not used because it only contained 82.6% complete BUSCOs. Our resulting *F. americana*, *F. nigra*, and *F. pennsylvanica* annotations contained 37,952, 36,821, and 37,079 gene models, constituting 96.7%, 97.3%, and 96.8% complete BUSCOs, respectively (Table 1), each over 10% more than the chromosome-level *F. pennsylvanica* assembly [17].

Our initial genome assemblies for *Fraxinus americana* and *F. pennsylvanica* did contain many uncollapsed haplotypic regions (Figure S2a, c, e), suggesting that our samples were highly heterozygous [28]. HapPy [28] indicated that the *F. americana*, *F. nigra*, and *F. pennsylvanica* assemblies were 75.8%, 99.6%, and 66.0% haploid, respectively (Table S4). This was corroborated by greater-than-expected gene model prediction numbers and duplicated BUSCO counts compared with the *F. pennsylvanica* reference in these initial assemblies, particularly for *F. americana* and *F. pennsylvanica* (Table S2). Lastly, self-vs-self frequency plots of synonymous substitution rates (Ks) between ostensibly syntenic duplicate gene pairs indicated uncollapsed haplotigs within the assemblies as peaks with low Ks that did not correspond to WGD events (Figure S3d, e, f).

We therefore generated largely haploid assemblies by post-processing using Purge Haplotigs [29], resulting in similarly high levels of complete BUSCOs of 96.0%, 95.1%, and 95.4% in *Fraxinus americana*, *F. nigra*, and *F. pennsylvanica*, respectively (Table 1). HapPy detected only haploid peaks for all three purged assemblies (Figure S2b, d, f), and self-vs-self Ks frequency plots between syntenic CDS pairs indicated that diploid sequences had largely been removed from these assemblies (Figure S4d, e, f). Still, our three *Fraxinus* assemblies as well as the Huff et al. [17] *F. pennsylvanica* assembly, had a high percentages of complete duplicated BUSCOs: 14.6%, 13.6%, 15.0%, and 13.9% for *F. americana*, *F. nigra*, *F. pennsylvanica*, and the *F. pennsylvanica* reference assembly, respectively. These high levels of duplicated BUSCOs may be a result of the relatively recent WGD in Oleeae. This pattern is also shared with close relatives, such as *Syringa oblata* [16], *Osmanthus fragrans* [10], and *Olea europaea* [14], which have 10.0%, 20.0%, and 13.3% duplicated BUSCOs, respectively. In the case of *O. fragrans* and *O. europaea*, these assemblies show unpurged heterozygosity in their self-vs-self SynMap Ks plots (Figure S5). Compared with the *F. pennsylvanica* reference assembly and the reference-guided *F. americana* and *F.*

*nigra* assemblies produced by Huff et al. [17], our assemblies contain slightly more heterozygosity in the form of low frequency $\log_{10}$ Ks values below − 1.0 between syntenic paralogous gene pairs. The Huff et al. [17] assemblies are virtually devoid of $\log_{10}$ Ks values below − 1.0 (Figure S4g, h, i).

Repeat annotation was executed again on the haplotig-purged assemblies. In all cases, our assemblies identified more Copia and Gypsy LTR elements, DNA transposons, and total interspersed repeats compared to the *F. pennsylvanica* reference assembly (Table S7). The assemblies from this study were also closer to genome sizes estimated by flow cytometry [22], with the reference-guided assemblies being over 100 MB smaller than estimated genome sizes. Again, these disparities may be partly due to better repeat detection in long-read genome assemblies, which likely contain fewer collapses of identical or near-identical sequences [24].

# Ploidy analysis

All three *Fraxinus* species sequenced for this study showed evidence of two polyploidy events following the gamma whole genome triplication (WGT-γ) at the base of all core eudicots (Figure S4d, e, f). When post-speciation and post-polyploidy gene retention was compared between syntenic blocks of *Fraxinus* and *Vitis vinifera*, which lacks any additional polyploidy events beyond WGT-γ [30], each *Fraxinus* species showed a six-to-one ratio against *V. vinifera* (Fig. 1b, c, d). Furthermore, the six syntenic blocks in *Fraxinus* exist in three groups of two, strongly suggesting an additional WGT event in the ancestry of the genus, followed by a single whole genome duplication (WGD) event. Previously, these (as presently known) Oleaceae-specific polyploidy events were proposed to be two or more WGDs [10, 17, 21, 31, 32]. Xu et al. [9] was the first to propose a secondary WGT at the base of Oleaceae, about 66 million years ago (Mya). Wang et al. [16] independently corroborated a WGT followed by a WGD for *Syringa oblata*. All currently published members of the tribe Oleeae (i.e., *Osmanthus fragrans* [11], *Olea europaea* [14], and *S. oblata* [16]) showed the same ploidy level as our *Fraxinus* species (Figure S6).

Wang et al. [16] also identified two subgenomes in *S. oblata*, the result of the most recent WGD event. The duplication that created these subgenomes was estimated to have occurred 28.71 Mya, very close to previous estimates [14, 17, 18, 21] and to the divergence of the *Syringa* genus from *Olea* and *Osmanthus*. The subgenomes had unequal chromosome numbers, with ten chromosomes in subgenome-A and thirteen in subgenome-B. All current Oleeae assemblies show strong macrosynteny with each subgenome of *S. oblata* [16], suggesting shared basic subgenome numbers of ten to thirteen chromosomes (Figure S6). All five Oleaceae tribes recognized by Wallander and Albert [33] have different basic chromosome numbers: $x = 23$ in Oleeae, $x = 11–13$ in Jasmineae, $x = 14$ in Forsythieae, $x = 13$ in Fontanesieae, and $x = 11$ Myxopyreae [34, 35]. The most thorough phylogenetic analysis to date suggests that Jasmineae and Forsythieae are most closely related to the Oleeae [36]. It seems likely that the ancestor of the Oleeae, Jasmineae, and Forsythieae had a basic chromosome number of thirteen or fourteen, with the ten chromosomes of subgenome-A being derived from a thirteen or fourteen-chromosome ancestor, perhaps via chromosome arm fusion.

*Jasminum sambac*, a more basal member of the Oleaceae, showed a three-to-one syntenic block ratio with *V. vinifera* (Figure S10), suggesting that it only contains the WGT event also seen in *Fraxinus*, corroborating the findings of Xu et al. [9]. *Forsythia suspensa* [12, 13] shows a one-to-one syntenic block ratio with *Jasminum* (Figure S11), suggesting that it also contains only the WGT since its divergence from *V. vinifera*. The exact details of the WGT lacks syntenic evidence outside of the Oleaceae, however, where it remains known only from that family. Xu et al. [9] placed the WGT at the base of the Oleaceae, but it is unclear if this event occurred before it diverged from its sister famiy, the Carlemanniaceae, as Zhang et al. [31] suggested. Zhang et al. detected an increase in duplicated ortholog groups in a clade containing *Silvianthus bracteatus* (Carlemanniaceae) and eight members of the Oleaceae, but without a published genome assembly to run syntenic analysis with, we will remain cautious in accepting evidence for a shared polyploidy event determined by synonymous substitution (Ks) peaks alone [37, 38].

When fractionation bias was calculated [39, 40] between each *Fraxinus* species and *Jasminum sambac*, two *Fraxinus* syntenic blocks displayed corresponding patterns of gene retention with one chromosome of each *J. sambac* (Figure S12), further indicating that there was indeed a WGD in the *Fraxinus* stem lineage since its divergence from *J. sambac*. Again, this pattern was the same between *Jasminum* and published genomes of other members of Oleeae. Taken together, these results support a WGT that includes at least the Oleeae, Jasmineae, and Forsythieae tribes, followed by a WGD in the ancestor of the Oleeae. Basic chromosome numbers for the Fontanesieae and Myxopyreae suggest the same ploidy level as in Jasmineae and Forsythieae, but syntenic data is lacking for them. Echoing the Ks evidence noted above, it is unclear if chromosome counts in the Carlemanniaceae ($x = 15$ in *Carlemannia* and $x = 19$ in *Silvianthus* [41]) reflect a shared WGT event with Oleaceae, as neither of these counts share the basic chromosome count with the Oleeae stem lineage, which appears to be 11 or 13 [34, 41].

# Population size history

Paleodemography for each *Fraxinus* assembly was determined using an R port for the Pairwise Sequentially Markovian Coalescent (PSMC) model [42]. Based on the reasonable assumption that variation within all three *Fraxinus* species should coalesce into the same, pre-speciation stem lineage, the demographic curve for *F. nigra* was shifted toward present day, also with smaller effective population size (Figure S13). This effect has been observed with simulated selfing data for different lychee (*Litchi chinensis* Sonn.; Sapindaceae) populations [43], suggesting that our *F. nigra* individual may stem from a more inbred population compared with *F. americana* and *F. pensylvanica*. By artificially lowering the mutation rate for the purposes of comparative analysis, *F. nigra* coalesced more closely with *F. amerciana* and *F. pennsylvanica* (Fig. 2). We also generated a PSMC curve for the *F. pennsylvanica* reference assembly using its own Illumina short-read data. When compared with the PSMCs for our three assemblies generated and analyzed using long-read data, the *F. pennsylvanica* reference individual showed a demography similar to our *F. nigra* individual (Figure S13), suggesting that it may also stem from a more inbred population compared with our own *F. pensylvanica* accession. The *F. pennsylvanica* from Huff at al. also lacked an upward “bump” in effective population size ($N_e$) that our *F. pennsylvanica* showed. As hypothesized for PSMC analyses in other species groups, this rise in $N_e$ may indicate a

change in population heterozygosity through an admixture event [44, 45]. Indeed, Huff et al. described their reference assembly as coming from an “admixed” population, rather than a “pure northern” or “pure southern” population [17]. As such, our *F. pennsylvanica* individual may come from a population with even more interspecies admixture than the Huff et al. reference accession, which may stem from a population with less admixture than previously thought. Also possible is that PSMCs generated using ONT long reads are less trustworthy indications of demographic size history, perhaps due to read accuracy differences, or that to the contrary, they actually reveal more runs of heterozygosity useful for demography estimation. When PSMCs were generated using Illumina short reads from the *F. pennsylvanica* reference mapped onto our *F. pennsylvanica* assembly instead, the $N_e$ hump is no longer visible (Figure S14a). Similarly, when our ONT long reads for *F. pennsylvanica* were mapped against the Huff et al. reference assembly, the $N_e$ hump is also missing (Figure S14b), indicating that there may be differences in population structure between our two individuals. However, it is also clear that use of homologous reference assemblies for PSMC analyses is an important factor, even infraspecifically, especially when assemblies are of different sizes and potentially capture different repetitive element profiles (e.g., [46]).

# Phylogenetic orthology inference

OrthoFinder was run on our assemblies and a selection of proteomes from other high quality Oleaceae, other Lamiales, other asterid, and rosid genome assemblies. A total of 703,440 genes (94.9% of total) were clustered into 34,447 orthogroups, 79,702 of which were species specific. An additional 7,604 orthogroups contained genes from all 21 species and 203 were single copy orthogroups.

The OrthoFinder species tree grouped the *Fraxinus* assemblies from Huff et al. properly with the assemblies generated in this study (Figure S15a). This tree differed from previous phylogenies [33, 34, 36] by not placing *Jasminum sambac* (Jasmineae tribe) sister to *Osmanthus*, *Olea*, *Syringa*, and *Fraxinus* (Oleaee tribe). Instead, *Forsythia suspensa* (Forsythieae tribe) was sister to the Oleeae, and *Jasminum* was sister to the rest of the Oleaceae. A recent study suggested that phylogenetic discordance in the placement of the five Oleaceae tribes is more likely due to hybridization and ancient widespread introgression, rather than incomplete lineage sorting (ILS) [36]. Furthermore, the Oleeae may have originated from a hybridization event between the Forsythieae and either the Jasmineae, or a “ghost lineage” that was sister to the Jasmineae. We also created a species tree using only orthogroups with one gene from each species (Figure S15b). This tree showed the same relationships as the OrthoFinder species tree that included orthogroups containing paralogs as well. The OrthoFinder gene duplication estimates at nodes (Fig. 3) agrees with the known polyploidy events, showing large increases in well-supported gene duplication events at the known and proposed WGT and WGD events.

# Conclusion

We constructed high quality genome assemblies for three critically endangered ash tree species: *Fraxinus americana* (white ash), *F. nigra* (black ash), and *F. pennsylvanica* (green ash). These assemblies had gene

space completeness comparable with the recent chromosome-level *F. pennsylvanica* reference assembly (Table 1). We demonstrate that using a single MinION flow cell can be sufficient for in-depth downstream analyses, such as synteny and ploidy evaluations, paleodemographic analyses, and phylogenomics. For example, our assemblies were sufficient to detect multiple polyploidy events in the evolutionary history of *Fraxinus.* Furthermore, this study was able to identify a six-to-one syntenic gene pair ratio between *Fraxinus* (as well as other published members of the Oleeae) against grapevine (Fig. 1). We also confirm that all sequenced members of the Oleaceae have undergone a WGT, corroborating the findings of Xu et al. [9] and Wang et al. [16], and that this event was followed by a subsequent WGD in the Oleeae tribe. Our study also showed that use of ONT long reads alone can permit detection of differences in population structure through PSMC analyses. We show evidence through PSMC analyses that our *F. nigra* individual comes from a more inbred population than our other *Fraxinus* samples (Figure S13), and correlate that finding with low levels of heterozygosity in the sample (Figure S2c). With further improvements in ONT error rates, long-read sequencing may have more utility in population structure analyses in the near future.

## Materials And Methods

# Material collection

Plant material and vouchers were collected from mature trees of unknown sex. *Fraxinus pennsylvanica* vouchers and material for DNA extraction were collected from Point Gratiot Park in Dunkirk, New York. *F. americana* and *F. nigra* vouchers and material for DNA extraction were collected from the College Lodge Forest in Brocton, New York (Table S5). Species identification was done in the field by coauthor Erik Danielson, Stewardship Coordinator with the Western New York Land Conservancy (https://www.wnylc.org/). The collections for material to extract DNA from took place on May 30, 2021, and vouchers from those same trees were collected on July 10, 2021. Voucher identifications were corroborated in the lab using Gleason and Cronquist [47] and Atha and Boom [48]. From each sample, about 10g of leaf tissue was placed into separate vials, flash frozen using liquid nitrogen, and stored at -80 C for later use.

# DNA extraction & sequencing

We followed the BioNano NIBuffer nuclei isolation protocol and the library preparation procedure from Pacific Biosciences – “Preparing Arabidopsis Genomic DNA for Size-Selected ~ 20 kb SMRTbellTM Libraries”. Additionally, Circulomics Short Read Eliminator (SRE) was used to remove reads less than 25kb in length with all samples. DNA sequencing was carried out on an Oxford Nanopore Technologies’ GridION instrument utilizing MinION flowcells (version R9.4). Genomic DNA libraries were prepared with the Oxford Nanopore Technologies ligation kit SQK-LSK110. Each flow cell underwent two washes in order to maximize the sequencing output. Sequenced reads were basecalled with the high-accuracy model in Guppy version 5.0.11. Read quality was assessed with Nanostat and NanoPlot [49] (Table S1, Figure S1).

# Genome assembly

The genomes for *Fraxinus americana* and *F. nigra* were assembled using Flye version 2.8.3 [50], and for *F. pennsylvanica*, Flye version 2.9 was used. All genomes were assembled using a minimum overlap between reads of 10,000 bp and the scaffolding option. The *F. nigra* assembly benefitted most from three of Flye's internal polishing iterations, while those of *F. americana* and *F. pennsylvanica* benefitted the most from one polishing iteration each. Mean genome size estimates were calculated using data from Whittemore et al. [22]. Assembly stats were generated using Quality Assessment Tool for Genome Assemblies (QUAST) version 5.0.2 [51], and assembly completeness was measured using Benchmarking Universal Single-Copy Orthologs (BUSCO) version 4.1.4 embryophyta_odb10 [23] (Table S2).

# Genome annotation

Initial annotations for the *Fraxinus* genome assemblies were generated using Gene Model Mapper (GeMoMa) version 1.7.1 [25, 26], a homology-based gene prediction software. Each GeMoMa annotation used the "Full Annotation" (Fraxinus_excelsior_38873_TGAC_v2.gff3) from *Fraxinus excelsior* BATG0.5 assembly [18] as a reference set. For the subsequent haploid and pseudoassemblies described below, additional references were used: *Olea europaea* (GWHAOPM00000000.gff) [14], *Jasminum sambac* (Js.gff) [9], and the *Arabidopsis thaliana* TAIR10 annotation that contains all genes (TAIR10_GFF3_genes.gff) [27]. Annotation completeness was measured using BUSCO version 4.1.4 embryophyta_odb10 [23] on the predicted proteins file output by GeMoMa (Table 1). BUSCO scores for GeMoMa annotations were corrected using GeMoMa's BUSCORecomputer. The *F. pennsylvanica* reference assembly was not used as a reference for GeMoMa because its extracted protein file had fewer complete embryophyta_odb10 BUSCOs. The *F. pennsylvanica* reference assembly contained only 82.5% complete BUSCOs, while *F. excelsior's* extracted proteins contained 92.7% complete BUSCOs. Proteins were extracted from the latter proteome using AGAT version 0.9.1 [52], and BUSCO scores were recomputed using Evidential Gene (http://arthropods.eugenes.org/EvidentialGene/).

*De novo* transposable element (TE) family identification and modeling was completed using RepeatModeler version 2.0.1 [53] and RepeatMasker version 4.0.1 [54] (Table S3). RepeatModeler Builddatabase and RepeatMasker both used the rmblast search engine (http://www.repeatmasker.org/RMBlast.html). In addition to the default RepeatScout [55] /RECON [56] pipeline, the LTR structural discovery pipeline was run as well for RepeatModeler.

# Syntenic Mapping, Ks Plots, Fractionation Bias, and Pseudoassemblies

Each genome assembly and annotation was uploaded to the Comparative Genomics (CoGe) online platform [57]. Each assembly was input into SynMap2 [58] and run against itself to identify polyploidy events. Syntenic dotplots and histograms of the synonymous substitution rates (Ks) between syntenic CDS pairs were generated for each self-vs-self comparison (Figure S3, S4a-f). Self-self plots were also created for *Osmanthus fragrans* (v1.0 id44014) and *Olea europaea* (v1.0 id63569) (Figure S5). Each

*Fraxinus* species was also run against *Vitis vinifera* (v12x, release 51 id62513; Figure S9) and the *F. pensylvanica* reference assembly (v1.4 id62886) to ascertain ploidy levels (Figure S8).

Pseudoassemblies were generated using SynMap's Syntenic Path Assembly option [59] between each *Fraxinus* assembly and the *F. pensylvanica* reference assembly [17]. These pseudoassemblies were annotated using GeMoMa as described above and reuploaded to CoGe. Duplicate subgenomes were assessed to determine relative ploidy level using SynMap's fractionation bias (FractBias) tool [40] [39]. Assembly and annotation statistics can be found in supplementary Table S6. Each pseudoassembly was compared against *Vitis vinifera* (v12x, release 51 id62513) [60] (Fig. 1b,c,d) and *Jasminum sambac* (v1.0 id62203) [9] (Figure S12b,c,d).

MCScan was utilized to generate syntenic dot plots, syntenic depth, and macrosynteny karyotype plots between assemblies (pseudoassemblies, for our *Fraxinus spp.*) [61]. Input annotation files were converted to gff3 format using AGAT version 0.9.1 [52]. AGAT was also used to extract CDS fasta files using each species' assembly and reformatted annotation. When detecting synteny between two species with the same ploidy level, a C-score cutoff of .99 (–cscore = .99) was used to filter out older duplication events for clearer connections between syntenic blocks. Otherwise, default options were used to generate Figures. GeMoMa [25, 26] annotations were also generated for *Forsythia suspensa* [12] and *Osmanthus fragrans* [10] using *Fraxinus excelsior*, *Olea europaea*, *Jasminum sambac*, and *Arabidopsis thaliana* as references because they lacked publicly available annotations.

# Measuring assembly diploidy

Diploidy of the initial *Fraxinus* assemblies was measured using HapPy (https://github.com/AntoineHo/HapPy). Each *Fraxinus* assembly was indexed using SAMtools version 1.14 [62]. ONT raw reads for each sample were mapped onto its corresponding assembly using Minimap2 version 2.20 [63]. Aside from not outputting secondary alignments, default options were used. The resulting BAM files were sorted and indexed using SAMtools. Each BAM file was used as input for the HapPy coverage command. Plots were generated using the HapPy estimate command. For *F. americana*, the lower threshold of coverage was 0, the upper threshold of coverage was 35, and the limit for the region between haploid and diploid peaks was 16 (Figure S2a). For *F. nigra*, the lower threshold of coverage was 0, the upper threshold of coverage was 35, and the limit for the region between haploid and diploid peaks was 1 (Figure S2c). For *F. pennsylvanica*, the lower threshold of coverage was 0, the upper threshold of coverage was 35, and the limit for the region between haploid and diploid peaks was 14 (Figure S2e). Haploidy was estimated to 75.78%, 99.57%, and 66.04% for *F. americana*, *F. nigra*, and *F. pennsylvanica*, respectively (Table S4).

# Purging heterozygosity

The partially diploid assemblies were reduced to haploid assemblies using Purge Haplotigs version 1.1.1 [29]. Each assembly was indexed using SAMtools version 1.11 [62]. The associated nanopore reads were mapped to each assembly using Minimap2 version 2.20 [63]. The resulting SAM file was sorted, converted to a BAM file, and indexed using SAMtools [62]. Each indexed and sorted BAM file was used as

an input for Purge Haplotigs. The outputs of repetitive elements and their positions in the genome assembly were converted into a BED-format file for input in Purge Haplotigs so those sequences would be ignored during the analysis.

Coverage histograms for each partially diploid assembly were created using Purge Haplotigs (Figure S7). Based on these plots, the low points between the haploid and diploid peaks were at read depths of 16, 2, and 17 for *F. americana*, *F. nigra*, and *F. pennsylvanica*, respectively. Diploid contigs were purged and genome stats were recalculated using QUAST [51] and BUSCO embryophyta_odb10 [23] (Table 1).

# Population size history

The reads for each assembly were mapped onto their corresponding assembly using Minimap2 version 2.20 [63]. Minimap2 used the Oxford Nanopore Technologies (ONT) preset and did not output secondary alignments. The SAM file was sorted, converted to a BAM file, and indexed using SAMtools version 1.14 [62]. SNPs were obtained using bcftools version 1.14 [64] mpileup with the ONT configuration and bcftools call using the original variant calling model and "-P0.01", outputting variant sites only. An R [65] port for the pairwise sequentially markovian coalescent (PSMC) model [42], PSMCR, was used to infer population size history (https://github.com/emmanuelparadis/psmcr). The number of PSMCR iterations was set to 30, the largest possible value for time to the most recent common ancestor (maxt) was set to 15, the atomic time interval pattern was "4 + 25*2 + 4 + 6", and 100 bootstrap replicates were calculated. A generation time (*g*) of 15 years and a mutation rate of 2.5 x $10^{-9}$ substitutions per site per year were chosen to match previous PSMC parameters used for *Fraxinus excelsior* [66]. For comparisons between the *F. pennsylvanica* generated in this study and the F. pennsylvanica reference assembly (figure S14), the only differences in this process were bcftools call using the multiallelic and rare-variant calling model and PSMCR maxt being set to 10.

PSMC curves of populations or species that share the same ancestral stem lineage, and so expected coalescence times and trajectories near those times, may differ in coalescence attributes because of inbreeding (i.e., heterozygosity) differences among accessions [43, 67]. Initial PSMC output provides the capacity to estimate pi (π; nucleotide diversity), where $\pi = 4N_e\mu$. From this formulation, nucleotide diversity, effective population size, and mutation rate are linearly related. From initial PSMC, *F. americana*, *F. nigra*, and *F. pennsylvanica* have 3,458,091, 2,143,896, and 3,928,012 segregating sites, respectively (these are the "sum_n" outputs in the .PSMC files, which we take to reflect heterozygosity). Since π roughly equals (number of segregating sites)/(genome size in basepairs), $\pi_{F.\ nigra} = 0.002763$ and $\pi_{F.\ pennsylvanica} = 0.004670$, etc., respectively. One can calculate a $\pi_{F.\ nigra}/\pi_{F.\ pennsylvanica}$ ratio to describe this difference, i.e., 0.591696, etc. We can then re-run PSMC using adjusted per-generation mutation rates to account for differences in heterozygosites. Of course, *g* could also be altered (as in Hu et al. [43]) from 15 years to reach a similar end. Using the $\pi_{F.\ nigra}/\pi_{F.\ pennsylvanica}$ ratios to adjust μ to 1.47924E-09 for *F. nigra* (etc.) aligns the curves – on the x-axis – to *F. pennsylvanica* at 2.5E-09, in years.

# Phylogenetic orthology inference

OrthoFinder [68, 69] was used to extract orthologous sequences from a group of Oleaceae, including our *F. americana*, *F. nigra*, and *F. pennsylvanica*, as well as *F. americana* (v0.2), *F. nigra* (v0.2), and *F. pennsylvanica* (v1.4) from Huff et al. [17], *Osmanthus fragrans* [10], *Olea europaea* [14], *Syringa oblata* [16], *Jasminum sambac* [9]), and *Forsythia suspensa* [12]. Additional Lamiales (*Callicarpa americana* [70], and *Tectona grandis* [71]), other asterids (*Coffea humblotiana* [72] and *Ophiorrhiza pumila* [73]), and rosids (*Vitis vinifera* [60], *Arabidopsis thaliana* [27], *Brassica rapa* [74], *Theobroma cacao* [75], *Prunus avium* [76], and *Populus tomentosa* [77]) were included. New GeMoMa annotations were generated for the *F. americana*, *F. nigra*, and *F. pennsylvanica* assemblies from Huff et al. because CDS and proteins could not be successfully extracted with the publicly available versions. As with the haploid and pseudoassemblies generated in this study, annotations were made using *Fraxinus excelsior*, *Olea europaea*, *Jasminum sambac*, and *Arabidopsis thaliana* annotations as references. These newly annotated versions were denoted by adding a “1” onto the end of the current version number. Each annotation was reduced to the longest isoform, and proteins were extracted using AGAT version 0.9.1 [52]. OrthoFinder was run with default settings. The species tree inference was performed with STAG [78], which uses the proportion of species trees derived from single-locus gene trees supporting each bipartition as its measure of support, and rooted using STRIDE [79]. Lastly, single-copy (“orthologs” only) gene trees were extracted as input for ASTRAL (version 5.7.4) species tree reconstruction [80]. In total, 203 trees were used as input for ASTRAL and default options were used.

# Declarations


## Acknowledgements

We thank Dan Fordham at Oxford Nanopore Technologies plc for the invitation to participate in the Org.One project. Additional support was provided by the National Science Foundation through awards DEB-2030871 to VAA and DEB-2139311 to CL.


## Data Availability

Raw reads for all species are available on ORG.one. *F. americana* is under the study PRJEB47186 and raw reads are samples ERS7485557, ERS7485558, and ERS7485568. *F. nigra* is under the study PRJEB47212 and raw reads are samples ERS7495131, ERS7495132, and ERS7495133. *F. pennsylvanica* is under the study PRJEB47234 and raw reads are samples ERS7495128, ERS7495129, and ERS7495130. All assembly assembly versions generated in this study are available on the Comparative Genomics (CoGe) online platform (https://genomevolution.org/coge/). Initial Flye assemblies: *F. americana* v1.0 id63747, *F. nigra* v1.0 id63746, *F. pennsylvanica* v1.0 id63751; haploid-purged assemblies: *F. americana* v3.0 id63752, *F. nigra* v3.0 id63759, *F. pennsylvanica* v3.0 id63760; Pseudoassemblies: *F. americana* v3.1 id63756, *F. nigra* v3.1 id63757, *F. pennsylvanica* v3.1 id63758; Reannotated Huff et al. assemblies: *F. americana* v0.21 id64423, *F. nigra* v0.21 id64424, *F. pennsylvanica* v1.41 id63910; Reannotated *Forsythia suspensa* v2.0 id64416; Reannotated *Osmanthus fragrans* v1.0 id64425.

# References


1. Jerome, D., et al. *Fraxinus americana*. The IUCN Red List of Threatened Species 2017 2017 March 23, 2022]; Available from: https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T61918430A61918432.en.
2. Jerome, D., et al. *Fraxinus nigra*. The IUCN Red List of Threatened Species 2017 2017 March 23, 2022]; Available from: https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T61918683A61918721.en.
3. Westwood, M., et al. *Fraxinus pennsylvanica*. The IUCN Red List of Threatened Species 2017 2017 March 23, 2022]; Available from: https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T61918934A61919002.en.
4. Westwood, M., et al. *Fraxinus profunda*. The IUCN Red List of Threatened Species 2017 2017 March 23, 2022]; Available from: https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T61919022A113525283.en.
5. Westwood, M., et al. *Fraxinus quadrangulata*. The IUCN Red List of Threatened Species 2017 2017 March 23, 2022]; Available from: https://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T61919112A61919114.en.
6. Herms, D.A. and D.G. McCullough, *Emerald ash borer invasion of North America: history, biology, ecology, impacts, and management.* Annual review of entomology, 2014. **59**: p. 13-30.
7. Kelly, L.J., et al., *Convergent molecular evolution among ash species resistant to the emerald ash borer.* Nature ecology & evolution, 2020. **4**(8): p. 1116-1128.
8. Klooster, W.S., et al., *Ash (Fraxinus spp.) mortality, regeneration, and seed bank dynamics in mixed hardwood forests following invasion by emerald ash borer (Agrilus planipennis).* Biological Invasions, 2014. **16**(4): p. 859-873.
9. Xu, S., et al., *A high-quality genome assembly of Jasminum sambac provides insight into floral trait formation and Oleaceae genome evolution.* Molecular ecology resources, 2022. **22**(2): p. 724-739.
10. Yang, X., et al., *The chromosome-level quality genome provides insights into the evolution of the biosynthesis genes for aroma compounds of Osmanthus fragrans.* Horticulture research, 2018. **5**.
11. Chen, H., et al., *Whole-genome resequencing of Osmanthus fragrans provides insights into flower color evolution.* Horticulture Research, 2021. **8**.
12. Li, L.-F., et al., *Genome sequencing and population genomics modeling provide insights into the local adaptation of weeping forsythia.* Horticulture research, 2020. **7**.
13. Li, Y., *Direct Submission*. 2022: College of Landscape and Art, Henan Agriculture University, Wenhua Road No. 95, Zhengzhou, Henan 450002, China.
14. Rao, G., et al., *De novo assembly of a new Olea europaea genome accession using nanopore sequencing.* Horticulture research, 2021. **8**.
15. Ma, B., et al., *Lilac (Syringa oblata) genome provides insights into its evolution and molecular mechanism of petal color change.* Communications biology, 2022. **5**(1): p. 1-13.
16. Wang, Y., et al., *A chromosome-level genome of Syringa oblata provides new insights into chromosome formation in Oleaceae and evolutionary history of lilacs.* The Plant Journal, 2022.

17. Huff, M., et al., *A high-quality reference genome for Fraxinus pennsylvanica for ash species restoration and research.* Molecular ecology resources, 2021.
18. Sollars, E.S., et al., *Genome sequence and genetic diversity of European ash trees.* Nature, 2017. **541**(7636): p. 212-216.
19. Alonge, M., et al., *Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing.* BioRxiv, 2021.
20. Landis, J.B., et al., *Impact of whole-genome duplication events on diversification rates in angiosperms.* American journal of botany, 2018. **105**(3): p. 348-363.
21. Unver, T., et al., *Genome of wild olive and the evolution of oil biosynthesis.* Proceedings of the National Academy of Sciences, 2017. **114**(44): p. E9413-E9422.
22. Whittemore, A.T., et al., *Ploidy variation in Fraxinus L.(Oleaceae) of eastern North America: Genome size diversity and taxonomy in a suddenly endangered genus.* International Journal of Plant Sciences, 2018. **179**(5): p. 377-389.
23. Manni, M., et al., *BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes.* arXiv preprint arXiv:2106.11799, 2021.
24. Wang, P., et al., *Impact of short-read sequencing on the misassembly of a plant genome.* BMC genomics, 2021. **22**(1): p. 1-18.
25. Keilwagen, J., et al., *Using intron position conservation for homology-based gene prediction.* Nucleic Acids Research, 2016. **44**(9): p. e89-e89.
26. Keilwagen, J., F. Hartung, and J. Grau, *GeMoMa: Homology-based gene prediction utilizing intron position conservation and RNA-seq data*, in *Gene Prediction*. 2019, Springer. p. 161-177.
27. *TAIR: The Arabidopsis Information Resource (TAIR)*. Available from: https://www.arabidopsis.org/download/index-auto.jsp%3Fdir%3D%252Fdownload_files%252FGenes%252FTAIR10_genome_release on www.arabidopsis.org [05/01/2022].
28. Guiglielmoni, N., et al., *Overcoming uncollapsed haplotypes in long-read assemblies of non-model organisms.* BMC bioinformatics, 2021. **22**(1): p. 1-23.
29. Roach, M.J., S.A. Schmidt, and A.R. Borneman, *Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies.* BMC bioinformatics, 2018. **19**(1): p. 1-10.
30. Chanderbali, A.S., et al., *Buxus and Tetracentron genomes help resolve eudicot genome history.* Nature communications, 2022. **13**(1): p. 1-10.
31. Zhang, C., et al., *Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications.* Molecular biology and evolution, 2020. **37**(11): p. 3188-3210.
32. Julca, I., et al., *Phylogenomics of the olive tree (Olea europaea) reveals the relative contribution of ancient allo-and autopolyploidization events.* BMC biology, 2018. **16**(1): p. 1-15.

33. Wallander, E. and V.A. Albert, *Phylogeny and classification of Oleaceae based on rps16 and trnL-F sequence data.* American Journal of botany, 2000. **87**(12): p. 1827-1841.
34. Dupin, J., et al., *Resolving the phylogeny of the olive family (Oleaceae): Confronting information from organellar and nuclear genomes.* Genes, 2020. **11**(12): p. 1508.
35. Taylor, H., *Cyto-taxonomy and phylogeny of the Oleaceae.* Brittonia, 1945. **5**(4): p. 337-367.
36. Dong, W., et al., *Phylogenomic approaches untangle early divergences and complex diversifications of the olive plant family.* BMC biology, 2022. **20**(1): p. 1-25.
37. Tiley, G.P., M.S. Barker, and J.G. Burleigh, *Assessing the performance of Ks plots for detecting ancient whole genome duplications.* Genome biology and evolution, 2018. **10**(11): p. 2882-2898.
38. Thomas, G.W., S.H. Ather, and M.W. Hahn, *Gene-tree reconciliation with MUL-trees to resolve polyploidy events.* Systematic biology, 2017. **66**(6): p. 1007-1018.
39. Tang, H., et al., *Screening synteny blocks in pairwise genome comparisons through integer programming.* BMC bioinformatics, 2011. **12**(1): p. 1-11.
40. Joyce, B.L., et al., *FractBias: a graphical tool for assessing fractionation bias following polyploidy.* Bioinformatics, 2017. **33**(4): p. 552-554.
41. Yang, X., S.-G. Lu, and H. Peng, *First report of chromosome numbers of the Carlemanniaceae (Lamiales).* Journal of Plant Research, 2007. **120**(6): p. 707-712.
42. Li, H. and R. Durbin, *Inference of human population history from individual whole-genome sequences.* Nature, 2011. **475**(7357): p. 493-496.
43. Hu, G., et al., *Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars.* Nature genetics, 2022: p. 1-11.
44. Walsh, J., et al., *Divergent selection and drift shape the genomes of two avian sister species spanning a saline–freshwater ecotone.* Ecology and evolution, 2019. **9**(23): p. 13477-13494.
45. Wang, C., et al., *Donkey genomes provide new insights into domestication and selection for coat color.* Nature communications, 2020. **11**(1): p. 1-15.
46. Patil, A.B. and N. Vijay, *Repetitive genomic regions and the inference of demographic history.* Heredity, 2021. **127**(2): p. 151-166.
47. Gleason, H.A.C.A., *Manual of vascular plants of northeastern United States and adjacent Canada*. 1991, Bronx, N.Y., USA: New York Botanical Garden.
48. Atha, D. and B. Boom, *Field Guide to the Ash Trees of Northeastern United States*. 2017, Bronx, NY: Center for Conservation Strategy, The New York Botanical Garden. 26 pp.
49. De Coster, W., et al., *NanoPack: visualizing and processing long-read sequencing data.* Bioinformatics, 2018. **34**(15): p. 2666-2669.
50. Kolmogorov, M., et al., *Assembly of long, error-prone reads using repeat graphs.* Nature biotechnology, 2019. **37**(5): p. 540.

51. Gurevich, A., et al., *QUAST: quality assessment tool for genome assemblies.* Bioinformatics, 2013. **29**(8): p. 1072-1075.
52. Dainat, J. *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (version 0.9.1)*. Available from: https://www.doi.org/10.5281/zenodo.3552717.
53. Smit, A. and R. Hubley, *RepeatModeler Open-1.0. 2008-2015.* 2015.
54. Smit, A., R. Hubley, and P. Green, *RepeatMasker Open-4.0. 2013–2015*. 2015.
55. Price, A.L., N.C. Jones, and P.A. Pevzner, *De novo identification of repeat families in large genomes.* Bioinformatics, 2005. **21**(suppl_1): p. i351-i358.
56. Bao, Z. and S.R. Eddy, *Automated de novo identification of repeat sequence families in sequenced genomes.* Genome research, 2002. **12**(8): p. 1269-1276.
57. Lyons, E. and M. Freeling, *How to usefully compare homologous plant genes and chromosomes as DNA sequences.* The Plant Journal, 2008. **53**(4): p. 661-673.
58. Haug-Baltzell, A., et al., *SynMap2 and SynMap3D: web-based whole-genome synteny browsers.* Bioinformatics, 2017. **33**(14): p. 2197-2198.
59. Lyons, E., et al., *Using genomic sequencing for classical genetics in E. coli K12.* PloS one, 2011. **6**(2): p. e16717.
60. Jaillon, O., et al., *The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla.* Nature, 2007. **449**(7161): p. 463-7.
61. Tang, H., et al., *Synteny and collinearity in plant genomes.* Science, 2008. **320**(5875): p. 486-488.
62. Li, H., et al., *The sequence alignment/map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-2079.
63. Li, H., *Minimap2: pairwise alignment for nucleotide sequences.* Bioinformatics, 2018. **34**(18): p. 3094-3100.
64. Danecek, P., et al., *Twelve years of SAMtools and BCFtools.* Gigascience, 2021. **10**(2): p. giab008.
65. Team", R.C., *R: A Language and Environment for Statistical Computing*, R.F.f.S. Computing, Editor. 2022.
66. Bai, W.N., et al., *Demographically idiosyncratic responses to climate change and rapid Pleistocene diversification of the walnut genus Juglans (Juglandaceae) revealed by whole-genome sequences.* New Phytologist, 2018. **217**(4): p. 1726-1736.
67. Low, Y.W., et al., *Genomic insights into rapid speciation within the world's largest tree genus Syzygium.* Nature communications, 2022. **13**(1): p. 1-15.
68. Emms, D.M. and S. Kelly, *OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy.* Genome biology, 2015. **16**(1): p. 1-14.
69. Emms, D.M. and S. Kelly, *OrthoFinder: phylogenetic orthology inference for comparative genomics.* Genome biology, 2019. **20**(1): p. 1-14.
70. Hamilton, J.P., et al., *Generation of a chromosome-scale genome assembly of the insect-repellent terpenoid-producing Lamiaceae species, Callicarpa americana.* GigaScience, 2020. **9**(9): p. giaa093.

71. Zhao, D., et al., *A chromosomal-scale genome assembly of Tectona grandis reveals the importance of tandem gene duplication and enables discovery of genes in natural product biosynthetic pathways.* Gigascience, 2019. **8**(3): p. giz005.
72. Raharimalala, N., et al., *The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of Coffea humblotiana, a wild species from Comoro archipelago.* Scientific reports, 2021. **11**(1): p. 1-14.
73. Rai, A., et al., *Chromosome-level genome assembly of Ophiorrhiza pumila reveals the evolution of camptothecin biosynthesis.* Nature communications, 2021. **12**(1): p. 1-19.
74. Belser, C., et al., *Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps.* Nature plants, 2018. **4**(11): p. 879-887.
75. Argout, X., et al., *The genome of Theobroma cacao.* Nature genetics, 2011. **43**(2): p. 101-108.
76. Tan, Q., et al., *Chromosome-level genome assemblies of five Prunus species and genome-wide association studies for key agronomic traits in peach.* Horticulture research, 2021. **8**.
77. An, X., et al., *High quality haplotype-resolved genome assemblies of Populus tomentosa Carr., a stabilized interspecific hybrid species widespread in Asia.* Molecular Ecology Resources, 2022. **22**(2): p. 786-802.
78. Emms, D. and S. Kelly, *STAG: Species tree inference from all genes. bioRxiv 267914*. 2018.
79. Emms, D.M. and S. Kelly, *STRIDE: species tree root inference from gene duplication events.* Molecular biology and evolution, 2017. **34**(12): p. 3267-3278.
80. Zhang, C., et al., *ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees.* BMC bioinformatics, 2018. **19**(6): p. 15-30.

# Tables

Table 1: Assembly and annotation statistics for *Fraxinus* species after running Purge Haplotigs. *F. pennsylvanica* v1.4 is the reference assembly from Huff et al.

| Assembly | *F. americana* | *F. nigra* | *F. pennsylvanica* | *F. pennsylvanica* v1.4 |
|---|---|---|---|---|
| # contigs | 4,364 | 2,544 | 6,764 | 110 |
| Largest contig | 3,688,385 | 6,590,250 | 6,486,905 | 56,547,140 |
| Est. Total length | 875 MB | 829 MB | 869 MB | 869 MB |
| Total length | 851,858,478 | 776,258,169 | 841,639,580 | 756,791,283 |
| GC (%) | 35.26 | 34.76 | 35.2 | 34.40% |
| N50 | 507,263 | 1,081,099 | 244,798 | 33,221,578 |
| L50 | 450 | 206 | 928 | 10 |
| # N's per 100 kbp | 0.8 | 0.48 | 0.62 | 12,120.86 |
| Complete BUSCOs | 1548 (96.0%) | 1536 (95.1%) | 1539 (95.4%) | 1565 (97.0%) |
| Complete single-copy BUSCOs | 1313 (81.4%) | 1316 (81.5%) | 1297 (80.4%) | 1341 (83.1%) |
| Complete duplicated BUSCOs | 235 (14.6%) | 220 (13.6%) | 242 (15.0%) | 224 (13.9%) |
| Fragmented BUSCOs | 29 (1.8%) | 33 (2.0%) | 32 (2.0%) | 26 (1.6%) |
| Missing BUSCOs | 37 (2.2%) | 45 (2.9%) | 43 (2.6%) | 23 (1.4%) |
| Annotation | *F. americana* | *F. nigra* | *F. pennsylvanica* | *F. pennsylvanica* v1.4 |
| gene model/mRNA count | 37,952/42,167 | 36,821/40,899 | 37,079/41,280 | 35,470/35,470 |
| Complete BUSCOs | 96.7% (1560) | 97.3% (1571) | 96.8% (1562) | 82.6% (1332) |
| Complete single-copy BUSCOs | 78.7% (1270) | 80.2% (1294) | 78.6% (1269) | 70.8% (1142) |
| Complete duplicated BUSCOs | 18.0% (290) | 17.2% (277) | 18.2% (293) | 11.8% (190) |
| Fragmented BUSCOs | 1.1% (17) | 0.9% (15) | 1.4% (23) | 2.4% (38) |
| Missing BUSCOs | 2.3% (37) | 1.7% (28) | 1.8% (29) | 15.0% (244) |

# Figures

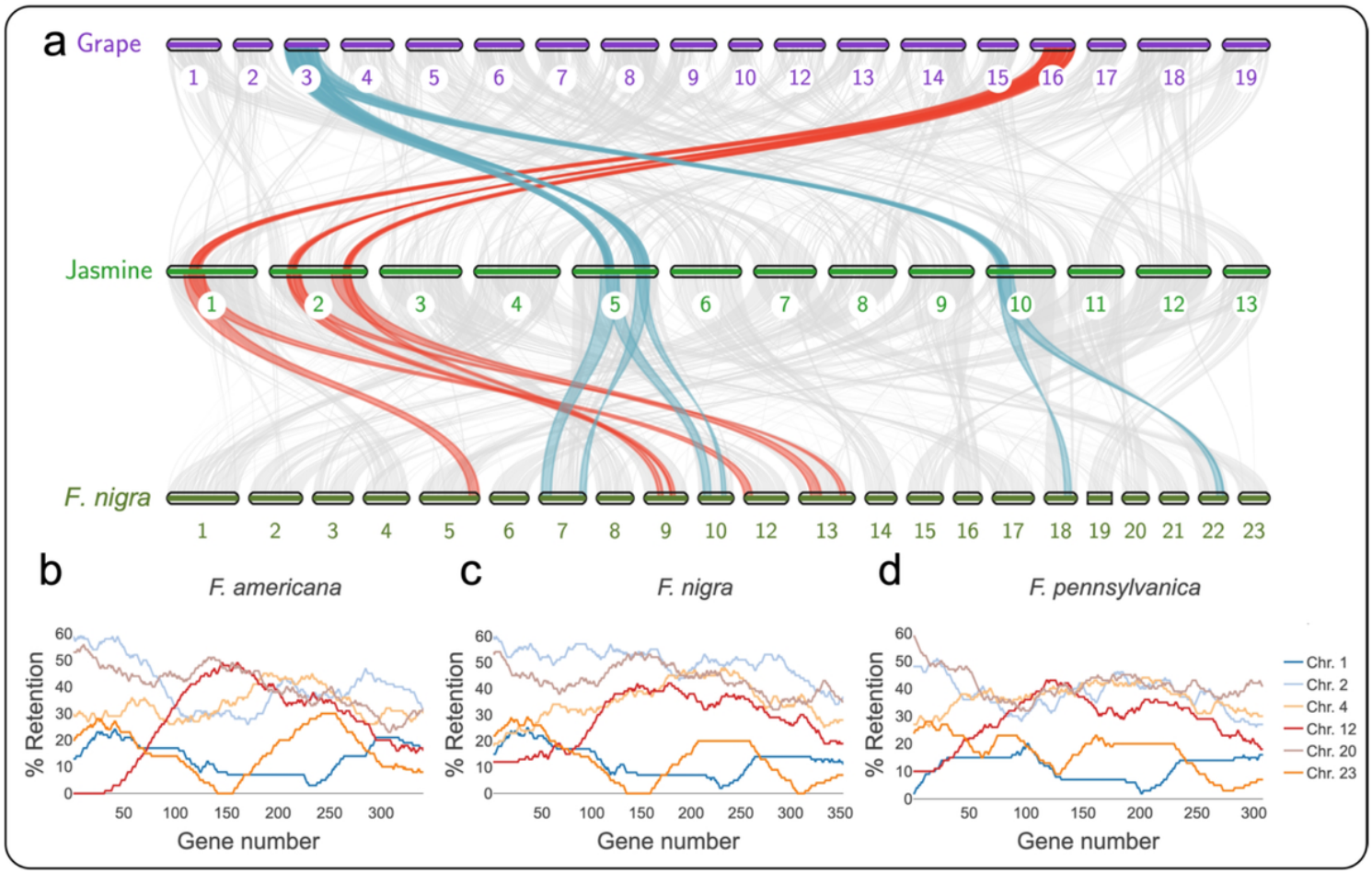



Figure 1

Polyploidy events in jasmine and ash since their divergence from grapevine. (a) Macrosynteny on a chromosome-scale between grapevine (*Vitis vinifera*) and jasmine (*Jasminum sambac*) and black ash (*Fraxinus nigra*). Numbers below chromosomes indicate chromosome number. Syntenic blocks between *Vitis*chromosome 3, *Jasminum,* and *Fraxinus* are connected by light blue lines. Syntenic blocks between *Vitis* chromosome 16, *Jasminum,* and *Fraxinus* are connected by orange lines. Fractionation bias between syntenic blocks of *Vitis* chromosome 19 and *F. americana* (b), *F. americana* (c), and *F. pennsylvanica* (d) reveals 3 pairs of subgenomes (for six subgenomes total), one pair of which retaining far fewer syntenic genes since the ancient WGT at the base of Oleaceae. The Y-axis represents the percent of gene retention of each *Fraxinus* syntenic block compared with windows of genes on *Vitis* chromosome 19. The X-axis represents 100-gene windows along *Vitis* chromosome 19. Line color corresponds with *Fraxinus* chromosome numbers in the key to the right of (d).

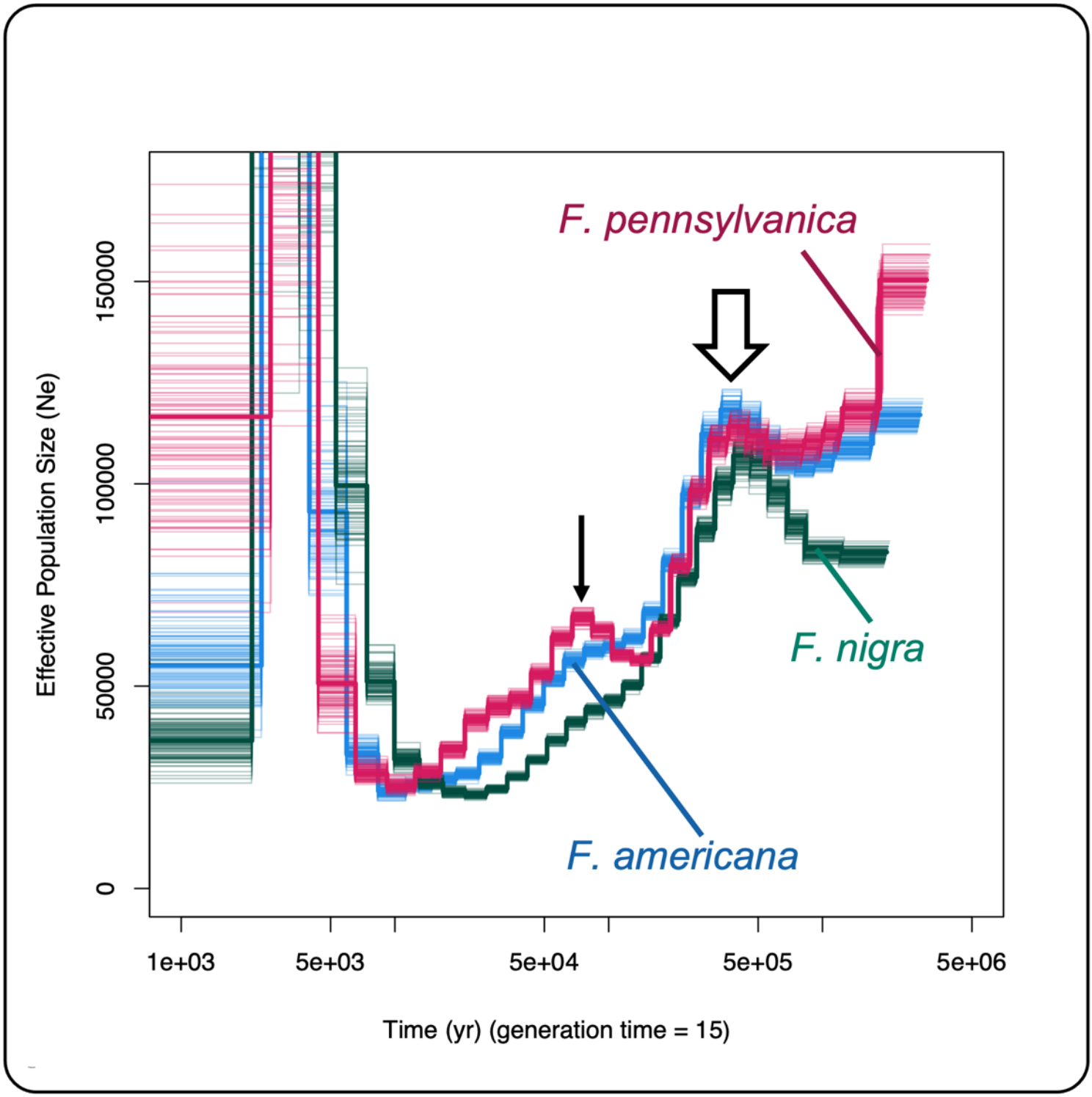
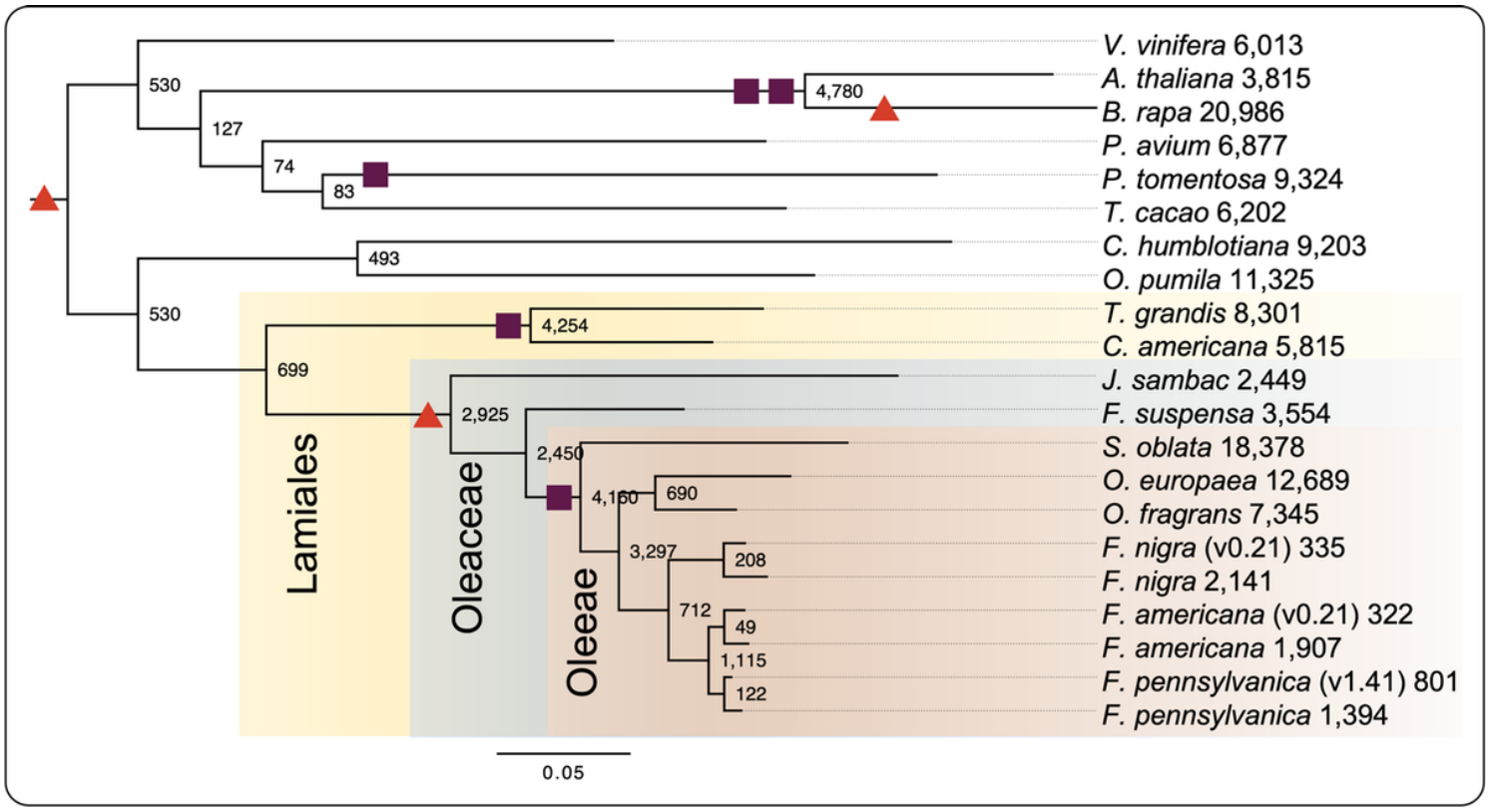



**Figure 2**

Pairwise Sequentially Markovian Coalescent with R (PSMCR) paleodemographic analyses for our three *Fraxinus* species. Demographic curve colors represent the following genome assemblies: blue: *F. americana*; green: *F. nigra*; pink: *F. pennsylvanica*. PSMCR plots generated by mapping *F. americana*, *F. nigra*, and *F. pennsylvanica* ONT long-reads against their own ONT long-read assemblies. All curves assume a generation time of 15 years. Mutation rates for *F. americana* and *F. nigra* curves have been corrected in order to coalesce similarly in ancient time with *F. pennsylvanica*. Mutation rates used were

2.1736x10$^{-9}$, 1.47924 x10$^{-9}$, and 2.5 x10$^{-9}$ nucleotide year$^{-1}$ for *F. americana*, *F. nigra*, *F. pennsylvanica*, respectively. The original bcftools variant calling model was used and tmax = 15 for PSMCR. The thin black arrow points to a pronounced Ne "hump" in *F. pennsylvanica*'s demographic curve. The hollow arrow points to a demographic feature, likely part of the *Fraxinus* stem lineage underlying all three species, wherein we infer the PSMCR curves should align.




## Figure 3

OrthoFinder species tree with inferred gene duplications and associated polyploidy events. Values at nodes and terminal branches are the numbers of gene duplication events with at least 50% of the descendant species maintaining both copies of the duplicated genes. Branch lengths represent average substitutions per site among orthogroup input trees. Red-orange triangles represent known or proposed whole genome triplication (WGT) events, with the triangle shared by all members of the tree being the ancient gamma hexaploidy event (WGT-g). Purple squares represent known or proposed whole genome duplication (WGD) events. The positioning of WGT and WGD events along branches is arbitrary and does not represent the relative timing of those events. The Lamiales order, Oleaceae family, and Oleeae tribe are highlighted and labeled. Tip labels are as follows: (Lamiales, Oleaceae) *F. pennsylvanica*: *Fraxinus pennsylvanica*; *F. pennsylvanica* (v1.41): *F. pennsylvanica* (Huff et al.); *F. americana*: *Fraxinus americana*; *F. americana* (v0.21): *Fraxinus americana* (Huff et al.); *F. nigra*: *Fraxinus nigra*; *F. nigra* (v0.21): *Fraxinus nigra* (Huff et al.); *O. fragrans*: *Osmanthus fragrans*; *O. europaea: Olea europaea*; *S. oblata*: *Syringia oblata*; *F. suspensa*: *Forsythia suspensa*; *J. sambac*: *Jasminum sambac*; (other Lamiales) *C. americana*: *Callicarpa americana*; *T. grandis*: *Tectona grandis*; (Gentianales) *C. humblotiana*: *Coffea humblotiana*; *O. pumila*: *Ophiorrhiza pumila*; (Rosids) *T. cacao*: *Theobroma cacao*; *P. tomentosa*: *Populus tomentosa*; *P.*

*avium*: *Prunus avium*; *B. rapa*: *Brassica rapa*; *A. thaliana*: *Arabidopsis thaliana*; *V. vinifera*: *Vitis vinifera*. Branch support values for the tree are shown independently in figure S15a; these are derived from the proportion of species trees derived from single-locus gene trees supporting each bipartition.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FraxinusGenomes120522supplementaryfigures.docx
- FraxinusGenomessuppltables.pdf