bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Title: 5S-IGS rDNA in wind-pollinated trees (*Fagus* L.) encapsulates 55 million years of reticulate evolution and hybrid origins of modern species

Authors: Simone Cardoni[1], Roberta Piredda[2], Thomas Denk[3], Guido W. Grimm[4], Aristotelis C. Papageorgiou[5], Ernst-Detlef Schulze[6], Anna Scoppola[1], Parvin Salehi Shanjani[7], Yoshihisa Suyama[8], Nobuhiro Tomaru[9], James R.P. Worth[10], Marco Cosimo Simeone[1]*

Affiliations

[1]Department of Agricultural and Forestry Science (DAFNE), Università degli studi della Tuscia, 01100 Viterbo, Italy

[2]Department of Veterinary Medicine, University of Bari, 70010 Valenzano, Italy

[3]Swedish Museum of Natural History, Stockholm, 10405 Sweden

[4]Unaffiliated, Orléans, France, ORCID ID: 0000-0003-0674-3553

[5]Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

[6]Max-Planck Institute for Biogeochemistry, 07701 Jena, Germany

[7]Natural Resources Gene Bank, Research Institute of Forests and Rangelands, Agricultural Research, Education and Extension Organization, Tehran, Iran

[8]Graduate School of Agricultural Science, Tohoku University, Osaki, Miyagi 989-6711 Japan

[9]Graduate School of Bioagricultural Sciences, Nagoya University, Chikusa-ku, Nagoya, 464-8601, Japan

[10]Ecological Genetics Laboratory, Forestry and Forest Products Research Institute (FFPRI), Tsukuba, Ibaraki, 305-8687 Japan

* **Corresponding author**

Marco Cosimo Simeone

Phone: +39 0761357352

Mail: mcsimeone@unitus.it

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Summary

Standard models of plant speciation assume strictly dichotomous genealogies in which a species, the ancestor, is replaced by two offspring species. The reality in wind-pollinated trees with long evolutionary histories is more complex: species evolve from other species through isolation when genetic drift exceeds gene flow; lineage mixing can give rise to new species (hybrid taxa such as nothospecies and allopolyploids). The multi-copy, potentially multi-locus 5S rDNA is one of few gene regions conserving signal from dichotomous and reticulate evolutionary processes down to the level of intra-genomic recombination. Therefore, it can provide unique insights into the dynamic speciation processes of lineages that diversified tens of millions of years ago. Here, we provide the first high-throughput sequencing (HTS) of the 5S intergenic spacers (5S-IGS) for a lineage of wind-pollinated subtropical to temperate trees, the *Fagus crenata – F. sylvatica* s.l. lineage, and its distant relative *F. japonica*. The observed 4,963 unique 5S-IGS variants reflect a complex history of hybrid origins, lineage sorting, mixing via secondary gene flow, and intra-genomic competition between two or more paralogous-homoeologous 5S rDNA lineages. We show that modern species are genetic mosaics and represent a striking case of ongoing reticulate evolution during the past 55 million years.

Keywords

Fagus, 5S rDNA, High-Throughput Sequencing, Wind-pollination, Adaptation, Hybrid evolution, Genetic diversity

Significance statement

The evolution of extra-tropical wind-pollinated tree genera involves dynamic speciation processes. High-throughput sequencing of the multi-copy, potentially multi-locus 5S rDNA reveals a complex history of hybrid origins, lineage sorting and mixing, and intra-genomic competition between paralogous-homeologous loci in the core group of Eurasian beech trees (genus *Fagus*) and their distant relative, *F. japonica*. The modern species are genetic mosaics and represent a striking case of at least 55 million years of ongoing reticulate evolution.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Introduction

Wind-pollination is a derived condition in most angiosperms and likely evolved in response to changing environmental conditions, with the grass family (Poaceae) being the most prominent example of a strictly wind-pollinated group (Friedman & Barrett, 2008). One of the most diverse and ecologically important groups of wind-pollinated woody angiosperms is the order Fagales including walnuts, birches, oaks and beech trees. Fagales lineages repeatedly shifted from biotic insect- to abiotic wind-pollination (Endress, 1977; Manos et al., 2001; Friis et al., 2011). Adaptations to wind-pollination were accompanied by spectacular changes in reproductive biology. Major innovations included spring flowering, the simplification of flowers and specialization of inflorescences, large stigmatic surfaces, lax male inflorescences, increase in stamen numbers, large anthers, high pollen production, and thin-walled pollen grains (Endress, 1977; Denk & Tekleva, 2014). The most striking innovation in the life history of many wind-pollinated Fagales groups is the delay of ovule development in relation to pollination (Endress, 1977). The absence of leaves facilitates pollination, whereas the beginning of the growing season favours fruit development. The general view is that the shift to wind-pollination in the Fagales promoted the group's rise to dominance in temperate forest ecosystems (Endress, 1977). However, how this costly change in reproductive biology affected speciation is not well understood (Friedman & Barrett, 2008).

Assessing processes involved in the origin and diversification of species is a central goal of taxonomy, ecology, biodiversity and phylogeny research (De Queiroz, 2007; Mallet, 2013; Scornavacca et al., 2020). In plants, these processes include cross-lineage genomic admixture, hybridization events and ancestral gene flow between allopatric and highly disjunct modern taxa (Willyard et al., 2009; Klein & Kadereit, 2016; Marques et al., 2016; Suarez-Gonzalez et al., 2016; Folk et al., 2017, 2018; McVay et al., 2017). Especially in trees, characterized by long generation times and large effective population size, wind-pollination acts as a natural catalyst for reticulate evolutionary processes such as chloroplast capture, introgression and increased interpopulation gene flow. Notably, plastid DNA variation is sorted by geography rather than speciation in all broadly sampled groups of Fagales studied so far (Fagaceae: Simeone et al., 2016, Yan et al., 2019; Nothofagaceae: Acosta & Premoli, 2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

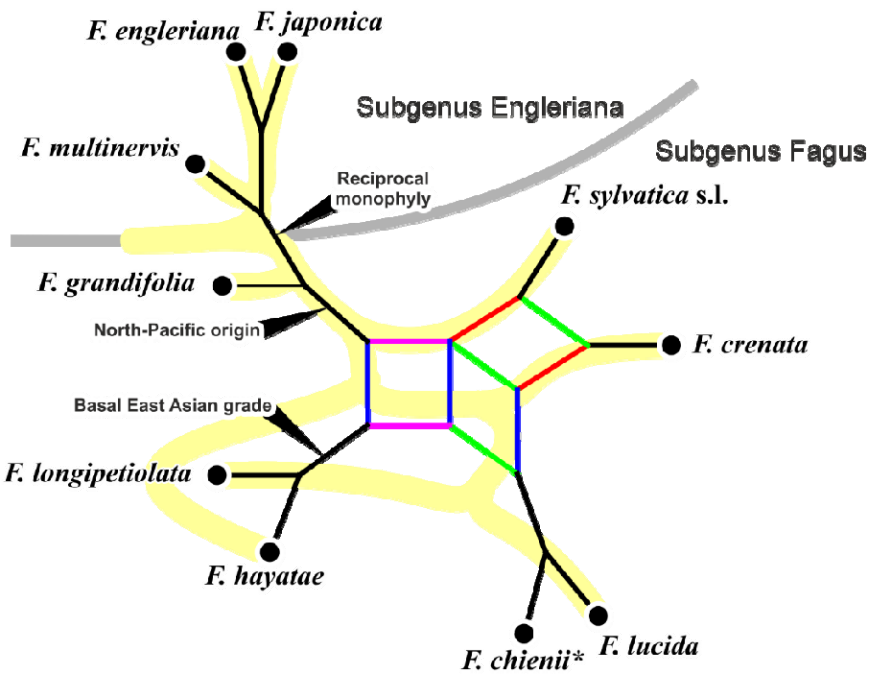


Fig. 1. Synoptical consensus network depicting known species relationships in *Fagus* (modified from Denk & Grimm, 2009, fig. 2). Alternative rooting scenarios (discussed in Denk & Grimm, 2009; Renner et al., 2016) indicated by black arrows. Black edges – non-conflicting phylogenetic splits; red – *crenata-sylvatica* lineage (within a clade including *F. lucida*, purple edge bundle) is supported by morphology/fossil record (Denk, 2003; Denk & Grimm, 2009), ITS data capturing substantial intra-individual polymorphism (Denk et al., 2005; Göker & Grimm, 2008), and when ITS individual consensus sequences are combined with *LEAFY* intron data (Renner et al., 2016); green – a closer relationship between *F. crenata* and *F. lucida* is indicated by intra-individual ITS polymorphic sequence patterns (Grimm et al., 2007); blue – a (semi-)continental East Asian clade (China, Taiwan) agrees well with morphology (Denk, 2003) and *LEAFY* intron data (Oh et al., 2016; Renner et al., 2016); yellow branches – species tree inferred by Jiang et al. (2021). *, there are no molecular data for *F. chienii* morphotypes; the species is probably a synonym of *F. lucida*.

Fagus L. (Fagaceae) provides an exemplary case. It is a small but ecologically and economically important genus of about ten monoecious tree species occurring in three isolated regions of the Northern Hemisphere: East Asia, western Eurasia, and eastern North America (Shen, 1992; Peters, 1997; Denk, 2003; Fang & Lechowicz, 2006). The genus probably originated at high latitudes during the Paleocene (western Greenland, northeast Asia; Fradkina et al., 2005; Denk & Grimm, 2009; Grímsson et al., 2016). It is subdivided into two informal subgenera (Shen, 1992; Denk et al., 2005). “Subgenus Engleriana” includes three (North-)East Asian species: *Fagus engleriana* Seemen ex Diels, *F. multinervis* Nakai, and *F. japonica* Maxim. “Subgenus Fagus” includes five or more Eurasian species: *Fagus sylvatica* L. s.l. (including *F. orientalis* Lipsky) in western Eurasia; *F. crenata* Blume in Japan; *F. lucida* Rehder & E.H.Wilson, *F. longipetiolata* Seemen and *F. hayatae* Palib. (incl. *F.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

pashanica C.C.Yang) in central and southern China and Taiwan, and a single North American species, *F. grandifolia* Ehrh. (including *F. mexicana* [Martínez] A.E.Murray). These two lineages diverged by the early Eocene (53 [62–43] Ma; Renner et al., 2016). While the lineage leading to the modern genus is at least 82–81 myrs old (Grímsson et al., 2016), extant species are the product of ~50 myrs of trans-continental range expansion and phases of fragmentation leading to diversification (Denk, 2004; Denk & Grimm, 2009; Renner et al., 2016). These dynamic migration and speciation histories left multifarious morphological and molecular imprints on modern members of the genus (Fig. 1).

Intra- and inter-specific phylogenetic relationships within *Fagus* have been difficult to resolve (Denk et al., 2002, 2005; Renner et al., 2016). In western Eurasia, for instance, where beech is the dominant climax species in mid-altitude forests, the number and rank of several taxa is still controversial (e.g. Gömöry et al., 2018). Difficulties arise from a low inter-specific but high intra-specific morphological diversity (e.g. Denk, 1999a, b) and equally complex inter- and intra-specific genetic differentiation in both the nucleome and plastome of beeches (Denk et al., 2002; Okaura & Harada, 2002; Grimm et al., 2007; Gömöry & Paule, 2010; Hatziskakis et al., 2009; Lei et al., 2012; Zhang et al., 2013). Most recently, Jiang et al. (2021) combined the data of 28 nuclear genes to infer a fully resolved (sub-)species tree using reconstructed haplotypes. This tree poorly reflects the signal in their gene sample indicating multiple events of gene duplication, incomplete lineage sorting below and above the species-level, and gene flow via secondary contact. Accordingly, ancient and reiterated hybridization/ introgression events were inferred for *Fagus* based on all assembled morphological, fossil and molecular data, and with respect to its biogeographic history and ecology (Denk et al., 2005; Denk & Grimm, 2009; Oh et al., 2016; Renner et al., 2016). Thus, *Fagus* constitutes a perfect example to study genetic complexity in a wind-pollinated genus of dominant forest trees.

Nuclear ribosomal DNA spacers, organized in multi-copy and potentially multi-loci arrays (Symonová, 2011), have great potential to detect past and recent reticulation events (e.g. Grimm & Denk, 2008, 2010). For *Fagus sylvatica*, Ribeiro et al. (2011) found two paralogous 5S rDNA loci, suggesting chromosomal rearrangements during its evolution. Although not studied in *Fagus*, intra- and inter-array variation can thus be expected. Availability of High-Throughput Sequencing (HTS) approaches is now boosting new efforts into research questions that were previously considered too expensive and labour-intensive (Babik et al., 2009; Glenn, 2011). In a recent study, Piredda et al. (2021) generated HTS amplicon data of

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

the non-transcribed intergenic spacers of the 5S rDNA in *Quercus* (Fagaceae) and demonstrated the great value of this approach to provide phylogenetic information for inspecting range-wide evolutionary patterns while circumventing previous technical limitations (e.g. cloning, abundant samples, special methodological frameworks; Denk & Grimm, 2010; Simeone et al., 2018).

In the present study, we explored the applicability of the same experimental pathway in six geographic samples of the core group of Subgenus Fagus: the *F. crenata* – *F. sylvatica* s.l. lineage, a dominant element of temperate mesophytic forests of Eurasia. The two species involved were resolved as sister species in previous phylogenetic studies (Denk et al. 2005; Renner et al., 2016; but see Fig. 1); they share a number of morphological key features such as leaf-like cupule appendages and rhombic-crenate leaves, and their modern distribution is highly disjunct at the western and eastern margins of the Eurasian continent. Hence, they are ideal to investigate evolutionary and speciation processes in wind-pollinated forest-forming trees. Our objectives were to test the utility of 5S-IGS HTS data for delineating beech species, assessing intra- and inter-specific diversity, and gaining deeper insights into an evolutionary history that involved reticulation. Further, we discuss differences in speciation processes and success in wind- and insect-pollinated trees.

Results

Pre-processing and removal of rare HTS reads (abundance < 4) yielded 145,643 sequences, corresponding to 4,693 representative 5S-IGS variants (work details, statistics and sequence structural features reported in Data S1, S2). Of these, 686 had an abundance ≥ 25 and were included in the phylogenetic analyses leaving 4007 variants to be classified using the Evolutionary Placement Algorithm (Data S2, S3). Analyses of 38 selected sequences further clarified the evolutionary relationships among the detected 5S-IGS lineages (Data S4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

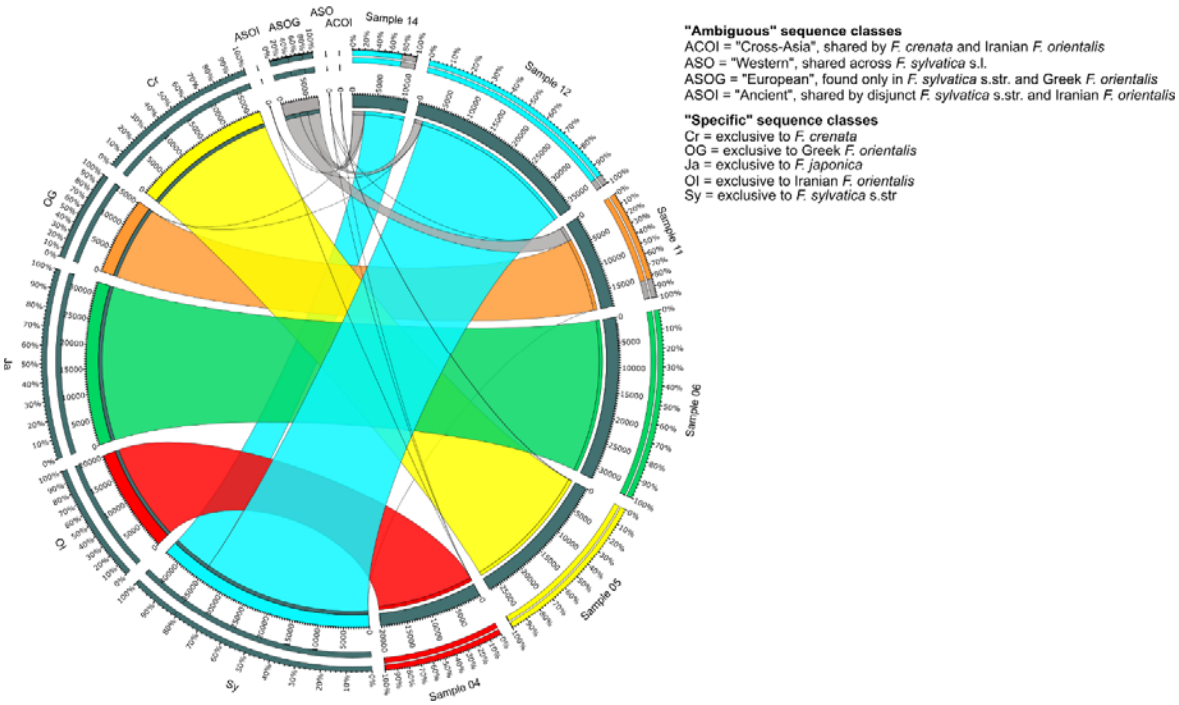


Fig. 2. Map of 5S-IGS sequence class versus samples (DNA extraction pools). Class "Ambiguous" refers to unique sequence variants shared between samples of different taxonomic identity (Iranian and Greek *F. orientalis* treated as different taxa); class "Specific" refers to variants exclusively found in a single taxon sample (or two samples, in the case of *F. sylvatica* s.str.).

686-tip 5S-IGS backbone phylogeny

We first categorized the 4,693 5S-IGS sequences into nine sequence classes based only on their sample distribution (Fig. 2). Variants were labelled as "specific" when exclusively found in one (or two, in case of *F. sylvatica* s.str.) sample of a specific taxon: "japonica", "crenata", "Iranian orientalis", "Greek orientalis", and "sylvatica". In addition, we identified four "ambiguous" classes, i.e. sequences shared among different taxa.

In a second step, five main 5S-IGS lineages (A, B, X, I, and O) were defined based on the phylogenetic analysis of 686 sequences with abundance ≥ 25 (Fig. 3).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

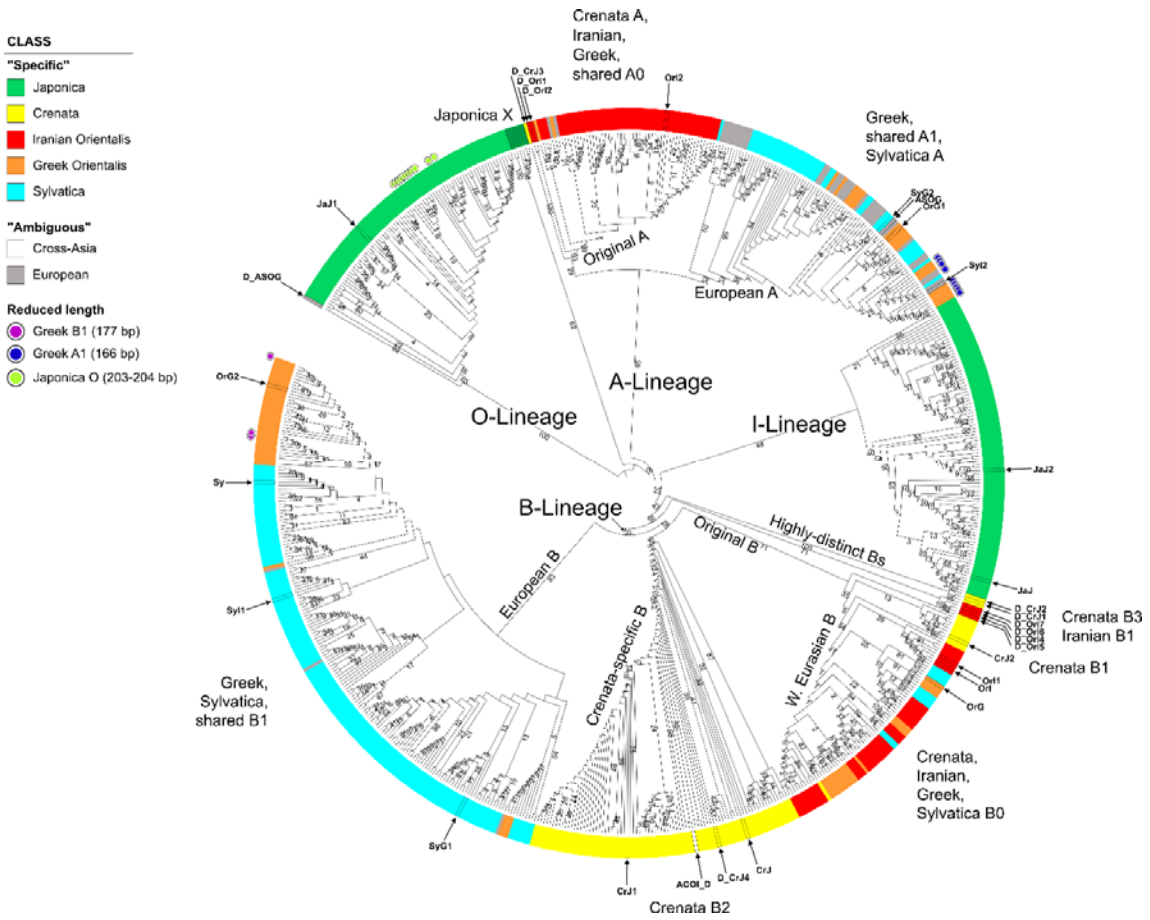


Fig. 3. Circular cladogram based on maximum likelihood (ML) tree inference and bootstrap (BS) analysis (550 BS pseudoreplicates) of the 686-sequence matrix including all 5S-IGS variants with a total abundance ≥ 25. The tree is rooted on the genetically most distinct O-type lineage; numbers at branches give ML-BS support. Colours indicate sequence class based on their sample distribution (Fig. 2); very short variants and sequences selected for the 38-sequence matrix (cf. Fig. 5) are highlighted.

The two ingroup 5S-IGS main types, 'A-Lineage' and 'B-Lineage', formed distinct clades (bootstrap [BS] support of 95/47) predating speciation within the *F. crenata – F. sylvatica* s.l. lineage (short: *crenata-sylvatica* lineage).

The Maximum Likelihood (ML) tree and Neighbour-Net (NNet) network (Fig. 4) show the duplicity of the five "specific" sequence classes. Most "ambiguous" variants are part of the 'European A' clade (BS = 34), while Iranian A-Lineage variants are exclusively found in the sister clade ('Original A'; BS = 53, BS = 29 when including 'Crenata A' variant). The lower support for the B-root (Fig. 3) relates to the higher diversity in the B-Lineage (Fig. 4; Table 1), which comprises two genetically coherent but diverse clades ('European B' and 'Crenata B2') and a poorly sorted clade 'Original B'. The ancestor of the *crenata-sylvatica* lineage was

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

polymorphic and the modern pattern affected by incomplete lineage sorting. The Iranian *F. orientalis* individuals represent a now genetically isolated sub-sample of the original variation found in the western range of the *crenata-sylvatica* lineage. In contrast, the Greek *F. orientalis* + *F. sylvatica* s.str. and *F. crenata* are better sorted.

Table 1 | Mean intra- and inter-type 5S-IGS divergence estimates (standard error shown above the diagonal). N = sample size (number of variants)

Type	N	Intra-/	Inter-type				
			A	B	I	X	O
A	958	0,059		0,014	0,015	0,016	0,021
B	2791	0,080	0,131		0,013	0,015	0,021
I	386	0,070	0,132	0,127		0,015	0,022
X	26	0,046	0,135	0,130	0,123		0,021
O	508	0,112	0,250	0,255	0,252	0,241	

Our data do not show ongoing lineage mixing in Japan between *F. japonica* (Subgenus Engleriana) and *F. crenata* (Subgenus Fagus). However, one of the *F. japonica* types, 'I-Lineage', is substantially more similar to both ingroup main types (A- and B-Lineage) than the other dominant *F. japonica* type ('O-Lineage'; Fig. 4). One ambiguous variant shared by western Eurasian beeches ('European O') is part of the O-Lineage; a small clade of *F. japonica*-exclusive variants ('Japonica X') nests between the A-B-I and the O-Lineage clades. *Fagus crenata* shares a relatively recent common origin with the western Eurasian beeches, represented by the 'Original B'-type (BS = 71) with two subgroups. Its 'Western Eurasian B' subclade (BS = 54; including one *F. crenata* B-variant: 'Crenata B0') is poorly sorted. Potentially ancient variants closest to 'Crenata B1' sequences (grade in Fig. 3; neighbourhood in Fig. 4) persisted in Italian, Greek and Iranian populations. In contrast to its western Eurasian relatives, *Fagus crenata* lacks non-deteriorated A-Lineage variants (Table 2; Data S1, section 4.2) with most variants falling within a highly supported, *F. crenata*-exclusive B-type clade ('Crenata B2'; BS = 98); this *F. crenata*-specific lineage also includes the 'Cross-Asia' shared variant found as a singleton in Iranian *F. orientalis*. The remainder ('Crenata B3') is placed between the I-Lineage and the core group of the B-lineage, together with sequentially highly derived Iranian B-type variants ('Iranian B1'). One sequence ('Crenata A') is placed within the 'Original A' clade, as sister to all other variants (Fig. 3; possibly a tree-branching artefact, cf. Fig. 4).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Table 2 | Main 5S-IGS sequence types observed and their evolutionary interpretation

Type	Taxonomic coverage	Evolutionary interpretation
O	*F. japonica*, co-dominant, mostly inconspicuous *Crenata-sylvatica* lineage, very rare, ± deteriorated [a]	Evidences ancient hybridization/ poly-ploidization event
I	*F. japonica*, co-dominant, inconspicuous	Reflecting shared ancestry with 'Subgenus Fagus'
A	*F. sylvatica* s.l., frequent, mostly inconspicuous *F. crenata*, very rare, deteriorated	Shared ancestral type, most diverse in Iranian *orientalis*
B	*Crenata-sylvatica* lineage, frequent, mostly inconspicuous	Lineage-specific types, diverse and ± sorted
Original B[b]	Relatively common, rare in *F. sylvatica* s.str.	Reflecting divergence between East Asia and western Eurasia
Iranian B1 Crenata B3	Iranian *F. orientalis*, uncommon, mostly inconspicuous *F. crenata*, uncommon, mostly inconspicuous	Strongly diverged types, reflecting an-cient genetic drift
Crenata B2 European B	*F. crenata*, most frequent, *inconspicuous* *F. sylvatica* s.str., Greek *F. orientalis*, *inconspicuous*	Species-specific types
Relict lineage	Western Eurasia, very rare; ± deteriorated	Reminiscent of ancient speciation proc-esses or reticulation
X	*F. japonica*, rare, inconspicuous	Unclear; potentially linked to non-sampled Chinese spp. of 'Subgenus Fagus'

[a] Signs of sequence deterioration include increased AT-content, accumulation of aberrant sequence features, and rare additional deletions.
[b] Two subtypes: Crenata B1, Western Eurasian B (incl. Crenata B0)

The only other subtype showing the same level of genetic coherence as ‘Crenata B2’ is the ‘European B’ type (BS = 93; Fig. 4), shared by *F. sylvatica* s.str. and Greek *F. orientalis* (Table 2). Given its distinctness, the ‘European B’ type might reflect most-recent sorting and speciation events that involved *F. sylvatica* and western (Greek) *F. orientalis* but not their eastern (Iranian) relatives. In general, the western Eurasian beeches form a genetic continuum characterized by several, partially incomplete sorting events (‘Iranian A’ vs. ‘European A’; ‘Original B’ vs. ‘European B’). Within this continuum, Iranian *F. orientalis* appears to be most isolated and ancient/ ancestral with respect to *F. crenata* and the I-Lineage of *F. japonica*.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

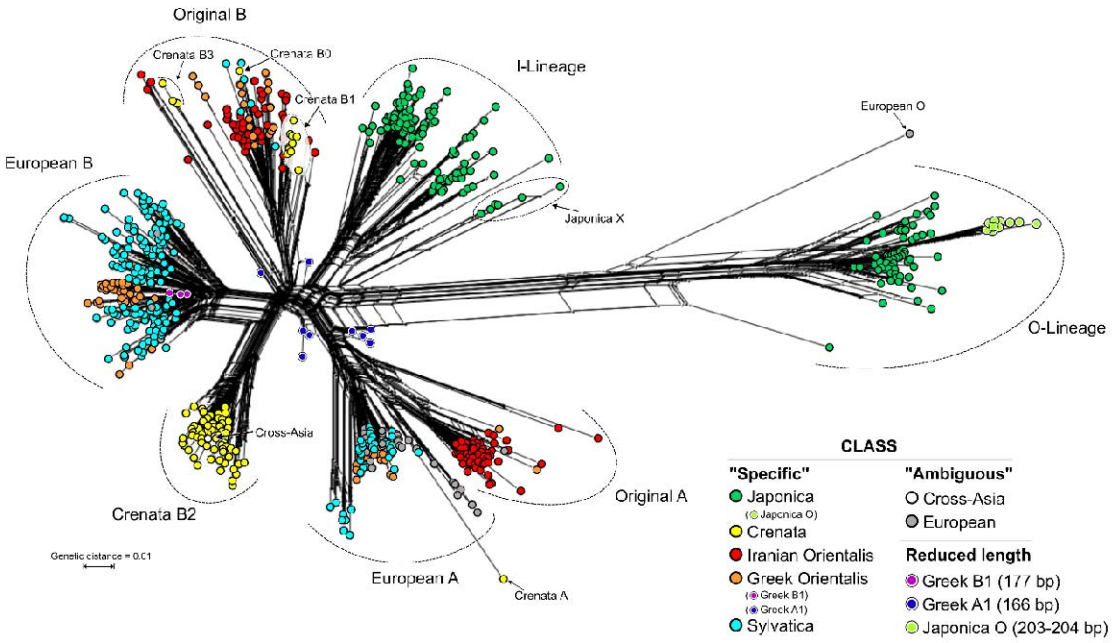


Fig. 4. Neighbour-net for the 686-sequence matrix, inferred from uncorrected (*p*-) pairwise distances. Neighbourhoods defined by well-defined interior "trunks" relate to prominent sorting events (bottleneck situations; evolutionary jumps) leading to coherent 5S-IGS lineages with high (near-unambiguous) root branch support in Fig.3; "fans" represent poor sorting of more ancient 5S-IGS variants forming incoherent lineages with ambiguous root branch support. Note the absolute genetic distance between the assumed outgroup, 'Japonica O'-type, and the ingroup types including the 'Japonica I' type. Colouration as in Fig. 3.

Very short 'Greek orientalis' variants are deeply embedded within the 'European A' and 'European B' clades; all short 'Japonica' variants are O-Lineage (Data S1, section 4.1). The distance-based NNet placed all short variants next to the centre of the graph (Fig. 4). Thus, they are sequentially undiagnostic lacking more than 100 bp from the 5' or central part of the spacer but also inconspicuous within the larger ingroup (*crenata-sylvatica* A-, B- and *F. japonica* I-Lineage).

Framework phylogeny

The ML trees for the 38 selected sequences representing the most abundant variants within each sample, strongly deviating variants, and "ambiguous" variants, show the deep split between outgroup (O-Lineage) and ingroup 5S-IGS variants (I-, A- and B-Lineage; Fig. 5). The mid-point root corresponds to this deep split, indicating an early (ancient) duplication of 5S rDNA arrays. Tip-pruning and elimination of the sequentially indiscriminative *F. japonica*

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

X-Lineage (Figs 3, 4; for details see Data S1) led to a marked increase in backbone branch support: the divergence between *F. japonica* ingroup variants (I-Lineage) and B-Lineage clades occurred after the isolation of the A-Lineage (BS = 99/84). A group of sequentially unique, rare (abundances < 25), partly shared variants form a distinct clade, a sixth lineage. This "Relict Lineage" is placed between the outgroup (O-Lineage) and ingroup subtrees (A-B-I clade). Detailed analysis of sequence structure (Data S1, section 4.3) shows that this placement is only partly due to potential ingroup-outgroup long-branch attraction. Both subtrees (O- and Relict Lineage) include variants most different from the ingroup (A-B-I-X) consensus, and branch support is higher when length-polymorphic regions are included. However, these rare shared types also have an increased number of mutational patterns, which appear to be primitive within the entire A-B-I-X lineage. Thus, they may represent relict variants; ancestral copies still found in 5S rDNA arrays that were subsequently replaced and eliminated within the *crenata-sylvatica* lineage (Table 2). Another important observation is that the long-branching A- and B-Lineage variants found exclusively in the Iranian sample comprise strongly divergent, genetically coherent lineages and two *F. crenata*-unique variants with no signs of pseudogeny. While some rare, shared ingroup variants (Relict Lineage types, 'Crenata A', some 'European O'; Fig. 4) show clear signs of sequence deterioration, others are inconspicuous (no pseudogenous mutations in flanking 5S rDNA) and can be highly similar to the most abundant variants.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

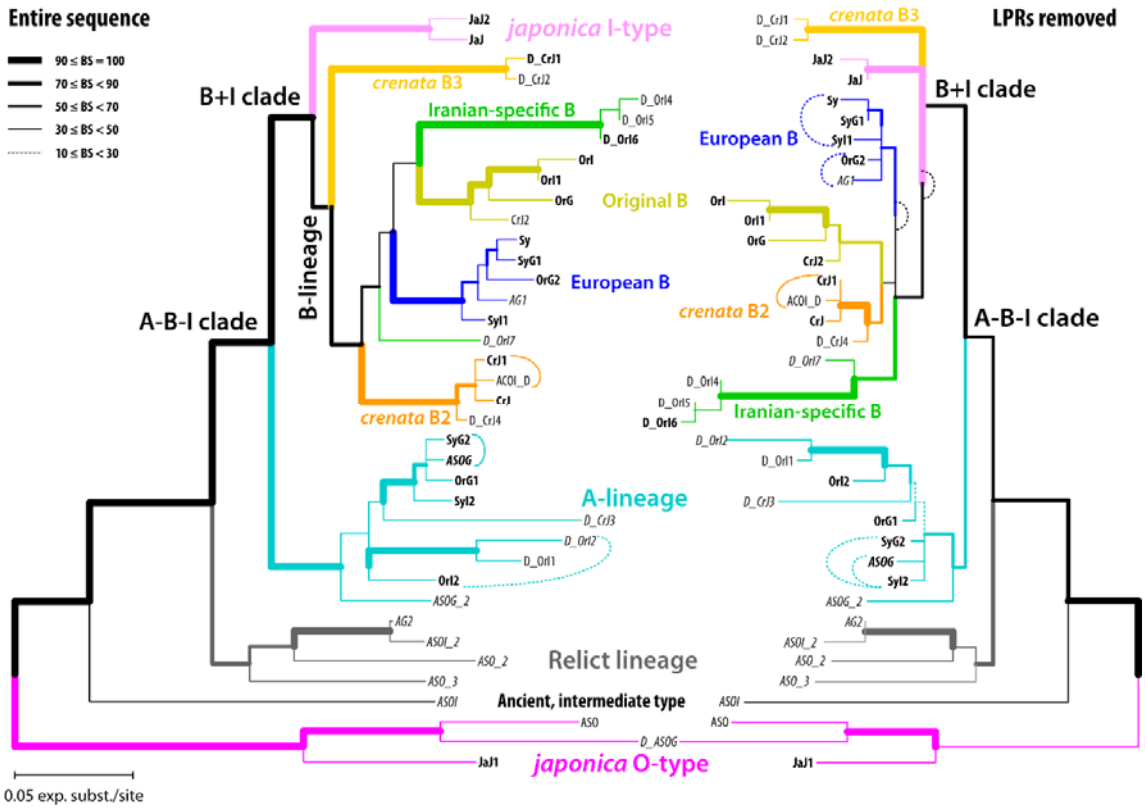


Fig. 5. Maximum likelihood phylograms inferred from the selected 38-sequence matrix including the most common variants of each main lineage, and rare, shared variants (labelled as "*A*…") as well as high-divergent variants ("*D*_…"). Left tree was inferred with generally length-polymorphic regions (LPR) included (as defined in Data S4, *Motives*); right tree with LPR excluded. Line thickness visualises non-parametric bootstrap (BS) support based on 10,000 BS pseudoreplicates.

Duplicity and differences between 5S-IGS populations in each sample

The unrooted, single-sample ML trees recovered two 5S rDNA clusters (5S main types) with high bootstrap values (71–100) in all studied samples (Fig. 6). This is in accordance with previous cytological results of Ribeiro et al. (2011), who found two functional 5S loci in (diploid) *F. sylvatica* s.str.. The main splits separate I- and O-Lineage variants in the outgroup *F. japonica* and A- and B-Lineage variants in *F. sylvatica* s.l. The difference between the two clusters is most pronounced in *F. japonica;* the least intra-species (intra-sample) divergence is exhibited by *F. crenata*, which largely lacks A-Lineage 5S-IGS. Phylogenetically intermediate variants characterize the western Eurasian samples (long-branched in Iranian *F. orientalis*); strongly modified variants with little affinity to either 5S rDNA cluster, hence, connected to the centre of the graph, are abundant in *F. crenata* (two lineages, one representing a relict 'Crenata A'-type; see Fig. 4) and *F. sylvatica* s.str. (a single lineage in

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

each population). These intermediate sequences do not show any structural peculiarity, except for the 'Crenata A'-type variant with a reduced GC content (34.3%; Data S1, section 4.2 and appendix B). The outgroup-type 'European O' represents the longest branches in the Greek *F. orientalis* (Greece) and both *F. sylvatica* s.str. samples (Fig. 6, samples 11, 12, 14). As a trend, the variants within the B- and I-Lineage show a higher divergence and phylogenetic structure than found within the A- and O-Lineage subtrees.

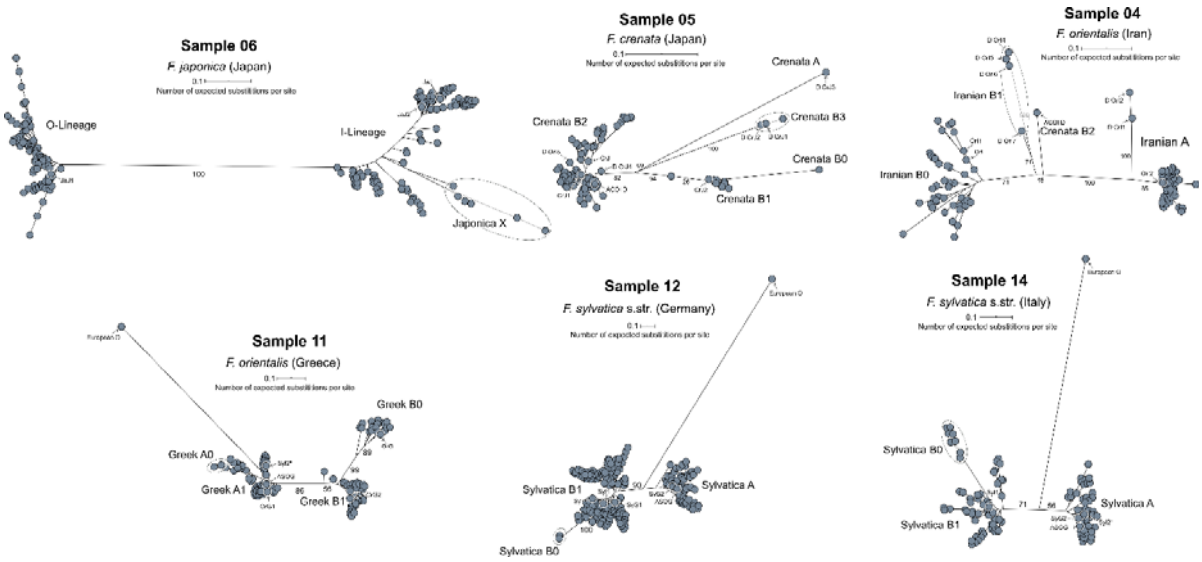


Fig. 6. Sample-wise maximum likelihood (ML) trees illustrating the bimodality of the 5S-IGS pool in each sample. Numbers give ML bootstrap support (based on 1000 pseudoreplicates) for selected major phylogenetic splits. Subtree and individual labels refer to lineages and tips introduced in Figs 3–5.

The Evolutionary Placement Algorithm (EPA) assessed the phylogenetic affinity of all variants with an abundance ≥ 4 not included in the 686-tip matrix (Data S2, S3). The A-/O- vs. B-/I-Lineage variants differ not only in GC content and sequence length but also relative abundance (Fig. 7). In the *crenata-sylvatica* lineage, the GC-richer B-Lineage variants are more frequent, while the GC-poorer A-Lineage sequences make up a higher portion of the HTS reads corresponding to rare sequence variants. The B-Lineage successively replaces the A-Lineage along a (south-)east (Iranian *F. orientalis*) to (north-)west (German *F. sylvatica*) gradient. In *F. japonica*, the ratio of I-Lineage, matching the B-Lineage in GC-content and length, to O-Lineage (GC-richest and longest variants) is ~2:1. In general, 5S-IGS arrays showed a high capacity to conserve signal from past reticulations and deep divergences: types

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

placed by EPA in 'European O' clade, representing a remnant, deteriorating sister lineage of the *F. japonica* O-Lineage (cf. Table 2; Figs 3, 4), can be found in all samples of the *crenata-sylvatica* lineage (Data S1, appendix A). In contrast, X-Lineage variants are exclusive to *F. japonica*. The single, low-abundant I-Lineage sequence identified by EPA in the *F. sylvatica* s.str. sample from Germany represents a relict variant from the initial radiation within the (A-)I-B clade. The Relict Lineage (cf. Fig. 5) is represented in all samples of the *crenata-sylvatica* lineage as well. Despite showing signs of beginning pseudogeny in the flanking 5S rRNA gene regions, its GC contents (35.1–40.0%) range between the median values of the O-Lineage and the I-B clade. Hence, they match the range seen in the A-Lineage, and are of the same length as most A- and B-Lineage types (Appendix B in Data S1).

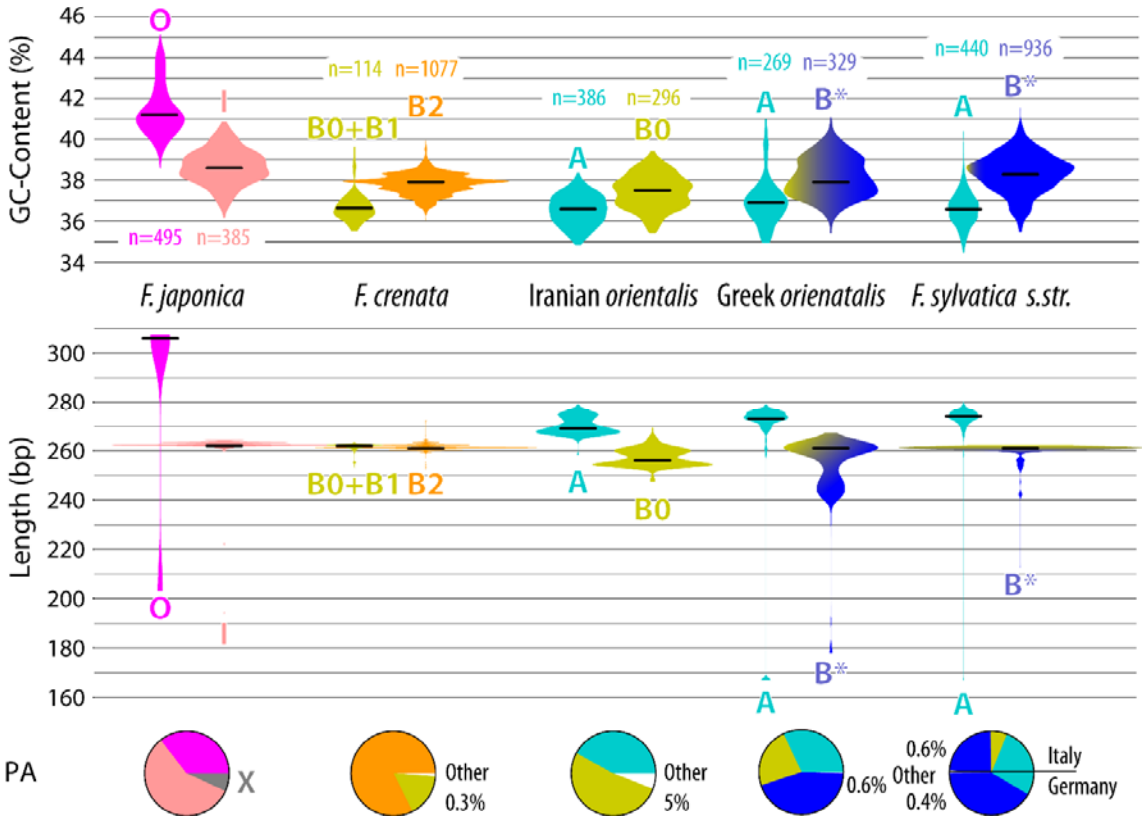


Fig. 7. Amplicon GC content and length violin plots for the (co-)dominant lineages/main 5S-IGS types found in each sample. Horizontal black lines give the median value for each main type (based on all variants of the respective lineage). Width of violin plots adjusted to visualise the relative proportion (number of HTS reads) of each type within a sample; *n* gives the plots' sample number (number of distinct variants). Pie-charts give the proportional abundance (PA) of the plotted types within each sample (see Data S1, appendix A, for absolute numbers). "O", "I", "A", "B0"–"B3" refer to respective (sub)types/ lineages labelled in the 686- and 38-guide trees (Figs 3, 5). Colouration gives affinity to main 5S-IGS lineages (Figs 3–5). *, excluding very rare Iranian *F. orientalis*- and *F. crenata*-specific B types (cf. Fig. 8).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Discussion

5S-IGS data to study complex history of wind-pollinated trees

Fagaceae have shifted at least twice from insect-pollination to wind-pollination during their evolutionary history leading to the rise to dominance of temperate tree genera such as *Fagus* and *Quercus*. These shifts left imprints in the genetic signature of the family recording ancient and recent gene flow between lineages that diverged tens of millions of years ago (e.g. McVay et al., 2017; López-Heredia et al., 2020). These characteristics may have contributed to the success story of these trees, but at the same time complicated phylogenetic reconstructions. The underlying processes can only be represented with a species network; any species tree (combined or coalescent, Jiang et al., 2021, or phylogenomic, Yang et al., 2020, using complete plastomes) will be incomprehensive, irrespective of the obtained support (Sections 4.5, 4.6 in Data S1). *Fagus* species are not only wind-pollinated but also animal-dispersed and have a narrow ecological niche (Peters, 1997). Therefore, beech populations are at risk of becoming isolated during phases of area disruption, which can lead to speciation accompanied by lineage sorting and increased genetic drift. At the same time, species boundaries remain permeable as expected in wind-pollinated plants. During phases of area expansion secondary contacts and lineage mixing occurred. In Subgenus Engleriana, up to three sequentially divergent ITS variants indicated (past) inter-species gene flow; likewise, *F. longipetiolata* and *F. pashanica* shared two distinct ITS lineages (Denk et al., 2005). Recently, the disparate gene trajectories seen in the polymorphic nuclear gene data of Jiang et al. (2021; Section 4.5 in Data S1; Data S5) provided further evidence for lineage mixing. Jiang et al.'s (2021) data indicate past, west-east introgression via the North Atlantic land bridge (cf. Grímsson & Denk, 2007) and suggest multiple secondary contacts shaping the gene pools of modern East Asian species of subgenus Fagus, most pronounced in the species pair *F. longipetiolata*/ *F. lucida* (Section 4.6 in Data S1). Hence, the objective of our study was to recruit data sets reflecting ancient, between ancestors of modern species, and recent, between modern species, reticulations.

The permeability of species boundaries sets apart wind- and insect-pollinated, common tree genera. A recent study on genetic consequences of pollination mode (Wessinger, 2021) showed that wind pollination leads to reduced population genetic structure. For example, in

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

one insect- and one wind-pollinated species of *Chamaedorea* (Arecaceae) in Mexico, the wind-pollinated species showed less isolation by distance as compared to the insect-pollinated species; the same was true in two common shrubs co-occurring in Mediterranean scrub, *Pistacia lentiscus* (wind-pollinated) and *Myrtus communis* (insect-pollinated). In addition, insect-pollinated, woody angiosperms with limited seed dispersal abundant in humid temperate regions such as *Acer* and *Rhododendron* are genetically coherent down to species-level (Grimm et al., 2006, 2007; Chappell et al., 2008; Grimm & Denk, 2014). Pairwise niche comparisons in temperate species of *Acer* have revealed only infrequent niche overlap continentally and globally (Grossman, 2021). In contrast, in the wind-pollinated *Fagus* displays neither niche specificity nor correlation with population sizes and occupied area (Cai et al., 2021). Genetic distances between sections of insect-pollinated *Acer* resemble those of subgenera or genera in wind-pollinated Fagaceae irrespective of the used marker (cf. various sequence data stored in gene banks). In some species (e.g. *Acer ibericum*; Grimm et al. 2007), intra-individual ITS variability can reach similar levels as found in individuals of *Fagus* (and exceeding variation in *Quercus*) but this variation is not shared among morphologically and (otherwise) genetically distinct species. Thus, in species of *Acer* both ITS and 5S-IGS variants, as far as studied, are highly species-diagnostic and sorted to a degree that allows the recognition even of cryptic species (Caucasian *A. orthocampestre*, Grimm & Denk, 2014) within a similar geographic range as western Eurasian *Fagus*. Furthermore, ITS and 5S-IGS data of *Fagus* converge to relationships incongruent with single-/low-copy nuclear genes (cf. Fig. 1; Data S5). In contrast, in *Acer*, the most recent nuclear-phylogenomic species tree (Li et al., 2019, using a 500-gene dataset) is largely congruent with earlier phylogenies using ITS/5S-IGS data/polymorphism. Likewise, species diversification rates in *Rhododendron* are more than ten times higher than in *Fagus* (Renner et al., 2016; Shrestha et al., 2018). The rise to dominance in temperate and boreal forest ecosystems of wind-pollinated Fagales taxa such as *Betula*, *Fagus*, and *Quercus*, appears to be linked with high within-population gene flow and permeable species boundaries. Consequently, the evolutionary history of such tree genera is different, more complex, and much more difficult to characterize than in insect-pollinated woody angiosperms.

The 5S intergenic spacers are a unique source of information to investigate complex evolutionary processes in wind-pollinated, temperate woody angiosperms. They are the only currently known region that (*i*) is high-divergent and multi-copy, (*ii*) can be easily amplified

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

due to the conservation of the 5S rRNA genes, and (*iii*) can be analysed using High-Throughput Sequencing because of its relatively short length. In this work, the O-Lineage and I/X-Lineage variants found in *F. japonica* indicate hybrid (allopolyploid) origin, and the distribution of A- and B-lineage variants in the *crenata-sylvatica* lineage reflect speciation processes predating the formation of the modern species. Hence, the 5S-IGS captures phylogenetic signal from 55 million years and multiple ancient hybridization events that are not (readily) traceable by other molecular markers (Oh et al., 2016; Yang et al., 2020; Jiang et al., 2021, but see Data S5).

Inter-species relationships and status of Iranian F. orientalis

Our data confirm the close relationship of *F. crenata* with *F. sylvatica* s.l. (including shared identical or highly similar 5S-IGS variants) and the deep split between the two Japanese species representing two subgeneric lineages (Shen 1992; Denk et al., 2005; Renner et al., 2016). The latter has recently been confirmed using single-/low-copy nuclear genes (Jiang et al., 2021). Jiang et al. (2021) suggested that East Asian species form a monophyletic group, setting apart the Japanese *F. crenata* and the western Eurasian *F. sylvatica* s.l. (Fig. 1). Not addressed in the original paper, Jiang et al.'s (2021) gene sample reveals a complex pattern of inter- and intra-species relationships in the East Asian species, partly involving western Eurasian and North American species (Data S5). While three genes (P21, P28, and P54) support an East Asian clade with BS values ≥ 53, nine of the 28 genes reject it. Instead, they support incompatible (conflicting) splits with BS values ranging from 61 (gene P38) to 99.7 (gene F138). The East Asian clade gets its support from the distinctness of western Eurasian and North American species and inconsistently shared, primitive or slightly evolved, genotypes (Fig. S18 in Data S1). Individuals carrying 'alien' gene variants (genes P4, P21, P49, P52, P69, and P72) furthermore point to pre- (ancient) or post-speciation (secondary) gene flow between *F. crenata* and *F. lucida* or *F. longipetiolata*, *F. hayatae* or *F. longipetiolata* and *F. lucida*, and Chinese *F. hayatae* (*F. pashanica*) and *F. longipetiolata.* Therefore, additional studies are needed to assess to which degree *F. crenata* and *F. japonica* 5S-IGS gene pools are impacted by primary or secondary contact with the predecessors of *F. hayatae*, *F. longipetiolata* and *F. lucida*. The mutation patterns seen in Jiang et al.'s (2021) data confirm, nonetheless, that the western Eurasian species (*F. sylvatica* s.l.) evolved from an ancestor genetically close to the present-day *F. crenata*. Although not resolved as sister clades

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

in trees based on single-/low-copy nuclear data (Oh et al., 2016; Jiang et al., 2021; but see Section 4.6 in Data S1), *F. crenata* and *F. sylvatica* s.l. are likely sister species.

Regarding the western Eurasian species, our results corroborate previous morphological (Denk, 1999a, b) and population-scale isoenzyme and genetic studies (Gömöry & Paule, 2010; Bijarpasi et al., 2020), which identified a split between disjunct populations traditionally considered as *F. orientalis* in Europe and adjacent Asia Minor, Iran, and the Caucasus. Furthermore, our data confirm the relatively recent contact and mixing between the westernmost *F. orientalis* (here represented by a Greek population) and *F. sylvatica* s.str. (Papageorgiou et al., 2008; Müller et al., 2019).

The Iranian *F. orientalis* must have been isolated for a much longer time and clearly represents a distinct species. According to Denk (1999b), Gömöry & Paule (2010) and Gömöry et al. (2018), the Iranian populations are morphologically and genetically distinct from both the western *F. orientalis* (SE. Bulgarian and NW. Turkish populations) and the Caucasian *F. orientalis*. The high amount of unique 5S-IGS variants can be explained by the geographic isolation and restriction of Iranian beech populations. However, they are not genetically impoverished (see also Erichsen et al., 2018). The overall diversity of 5S-IGS A- and B-Lineage variants points to a complex history of Iranian beech within the limits of its current distribution, characterized by strong topographic heterogeneity (sea-level to close to 3,000 m a.s.l.) and pronounced small and large-scale climatic gradients (Sagheb-Talebi et al., 2014). One possible further source are the Caucasian populations geographically closer to the Iranian, which may have been in contact in the more recent past (via Cis-Caucasia and Azerbaijan). Hence, some of the Iran-specific variants in our study may be shared with Caucasian populations. Additional data from the Caucasus and north-eastern Turkey are needed to quantify the degree of this contact and to clarify the taxonomic status of these populations.

Other genetic imprints in the Iranian populations might be a legacy from extinct beech populations growing to the east and possibly from the disjunct populations of the Nur (Amanos) Mts in south-central Turkey. The fact that *F. crenata* and Iranian *F. orientalis* still carry exclusively shared ('Cross-Asia Crenata B2') and related ('Crenata A'; 'Crenata B0') 5S-IGS variants (Figs 3, 4) fits with a more or less continuous range of beech populations from the southern Ural, via Kazakhstan and Siberia, to the Far East until the end of the middle Miocene (see *Formation of 5S-IGS gene pools*).

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

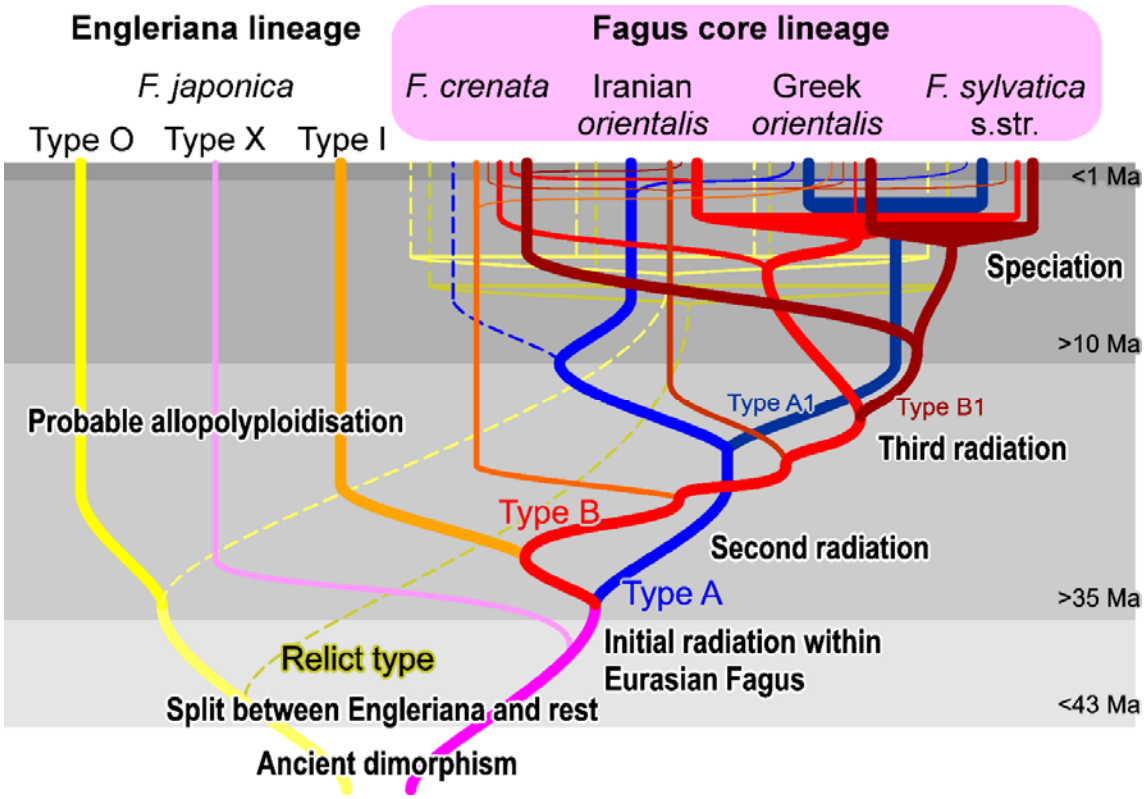


Fig. 8. Doodle summarising the totality of our results and their interpretation regarding all available (referenced in the text) information.

Ancient polyploidization and past reticulation in beech

The *Fagus crenata-sylvatica* s.l. lineage represents the most widespread, recently diverged and diversified branch of the genus (Data S4; Denk & Grimm, 2009; Renner et al., 2016). Differentiation patterns observed in the 5S-IGS provide a unique temporal window into the evolution of a wind-pollinated tree genus (Fig. 8): speciation and isolation led to the accumulation of new lineage-specific variants, which were then exchanged or propagated during episodes of favourable climate and massive expansion of beech forests.

New 5S-IGS variants would have been easily dispersed via pollen over long distances, while the genetic characteristics of the mother trees (passed on via ovules) remained geographically constrained. The main dispersal vector for *Fagus* seeds have been sedentary jay birds (Darley-Hill & Johnson, 1981), a group of corvids that evolved in the middle Miocene (Mayr, 2005) coinciding with the highest taxonomic diversity recorded in *Fagus* (Denk & Grimm, 2009). Another important animal seed-dispersal vector are small rodents, with specialized tree

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

squirrels occurring since the early Oligocene (Steppan et al., 2004; Casanovas-Vilar et al., 2018); a time when first major beech radiations took place (Denk & Grimm, 2009; Pavlyutkin et al., 2014).

Additional data from Chinese and Taiwanese species will be needed to assess whether the A- or B-Lineage, or the partly pseudogenic Relict Lineage, represent the original stock of western Eurasian beeches and their Japanese sister species. The A-Lineage is largely lost in *F. crenata* and sequentially equally distant to the B-Lineage and the *F. japonica* ingroup variants (I-Lineage; Fig. 4). Therefore, we can expect that both lineages, A and B+I (and possibly the X-Lineages) are present at least in some of the Chinese species and represent an ancient polymorphism shared by all Eurasian Subgenus Fagus species (Fig. 8; cf. Denk et al., 2005, for ITS). Likewise, our assessment of the 28-gene data of Jiang et al. (2021; Data S5) revealed poor genetic isolation of *F. crenata* from all other East Asian species of subgenus Fagus; confirmed by upcoming, densely sampled complete plastome data focussing on Japanese beech (first results accessible on ResearchGate: Worth et al., 2021). This is consistent with a main finding of the present study that *Fagus crenata* and its precursor(s) exchanged genetic material with other East Asian beeches when getting into contact.

The pronounced ancient nuclear polymorphism, also a predominant feature in Jiang et al.'s (2021) data, implies that modern beeches are of hybrid or allopolyploid origin. Our 5S-IGS data clearly indicate a hybrid origin for the species of Subgenus Engleriana, with the *F. japonica* I-Lineage variants representing common ancestry with Subgenus Fagus (Eurasian clade), while the outgroup variant (O-Lineage) represents another, potentially extinct, lineage of high-latitude beeches (Denk & Grimm, 2009; see also Fradkina et al., 2005; Grímsson et al., 2016). Hybrid/ allopolyploid origins would also explain the great divergence observed in the ITS region of *Fagus*, especially Subgenus Engleriana (Denk et al., 2005), a lineage with a poorly sorted nucleome and still sharing ± primitive genotypes with North American and East Asian members of Subgenus Fagus (Data S5). This hypothesis receives strong support from fossil evidence. Oldest unambiguous fossils of *Fagus*, dispersed pollen grains from the early Paleocene of Greenland, of small size and with narrow colpi reaching almost to the poles (Grímsson et al., 2016), resemble Subgenus Engleriana. The subsequent radiation of beeches involved western North America and East Asia, where fossil-taxa combining morphological characteristics of both modern subgenera are known (Denk & Grimm, 2009). For example, the early Eocene *F. langevinii* has branched cupule appendages as exclusively found in extant

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

F. engleriana along with Subgenus Engleriana type pollen but resembles Subgenus Fagus in features of leaf and nut morphology (Manchester & Dillhoff, 2005). Dispersed pollen from the late Eocene of South China also resembles modern pollen of Subgenus Engleriana (Hoffman et al., 2019). This might reflect an early phase in the evolution of *Fagus*, during which the modern subgenera were evolving, but not fully differentiated. Consequently, ancient genetic links can be traced in Jiang et al.'s (2021) data as well (Fig. S18, Tables S3, S4 in Data S1). By the early Oligocene (34 Ma), fossil-species can clearly be assigned to either Subgenus Fagus or Engleriana. *Fagus pacifica* from western North America resembles Subgenus Fagus in leaf architecture and the cupule/nut complex (Meyer & Manchester, 1997); cupules and leaves from the Russian Far East can unambiguously be ascribed to Subgenus Engleriana (Pavlyutkin et al., 2014; as *F. evenensis*, *F. palaeojaponica*, *Fagus* sp. 3).

Past reticulation is conceivable for the European species as well. The first beeches arrived in Europe in the Oligocene showing a general morphotype found across continental Eurasia (*F. castaneifolia*; Denk, 2004; Denk & Grimm, 2009). During the Miocene, *F. castaneifolia* was gradually replaced by *F. haidingeri*, the ancestor and precursor of all contemporary western Eurasian beeches (Denk et al., 2005; Denk & Grimm, 2009). Like *F. castaneifolia*, *F. haidingeri* shows high morphological plasticity. Hence, this fossil-species may represent a species complex rather than a single species. In addition, in southern Europe, gene flow between *F. haidingeri* and another fossil-species, *F. gussonii*, might have occurred. Jiang et al.'s (2021) genes F289, P14, P21, P52, and P54 document past, unilateral introgression via the now unavailable North Atlantic land bridge (Grímsson & Denk, 2007) from the North American into the western Eurasian lineage (Section 4.6 in Data S1). Our data confirm the potential for (sub)recent reticulation and incomplete lineage sorting between and within the morphologically distinct Greek *F. orientalis* and *F. sylvatica* s.str., and similar processes likely occurred repeatedly since the Miocene. Notably, the Greek *orientalis* and *F. sylvatica* s.str. comprise three main 5S-IGS lineages compared to only two in the Iranian sample. In addition to the inherited polymorphism shared with the Iranian *F. orientalis* (relatively similar 'Original A' and 'European A' types within the A-Lineage; shared 'Western Eurasian B'), we found a highly abundant, moderately evolved but less diverse B-Lineage variant (European B in Figs 3–6), exclusive to Greek *orientalis* and *F. sylvatica* s.str., likely reflecting the recent common origin of these species.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Spatio-temporal framework for gene pool formation in the F. crenata – F. sylvatica *s.str. lineage*

With currently available methods, it is impossible to date potentially paralogous-homoeologous multi-copy 5S IGS data. Nevertheless, the 5S IGS captured complex evolutionary signals, likely covering the past 55 million years. During the Eocene (56–33.9 Ma; Cohen et al., 2013, updated), the lineages leading to *F. japonica* and *F. crenata* + *F. sylvatica* s.str. started to diverge and speciation in the *crenata-sylvatica* lineage probably started in the late Oligocene (Chattian; 27.82–23.03 Ma; see above; Renner et al., 2016). The fossil record indicates that gene flow between the western(-most) and eastern(-most) populations of the *crenata-sylvatica* lineage became impossible after the middle Miocene, when the near-continuous northern Eurasian distribution of beech forests became fragmented at around 15–10 Ma (Denk, 2004; Arkhipov et al., 2005). Because of their geographic closeness, eastern (NE./E. Asian) populations of the *crenata-sylvatica* lineage remained in closer and longer contact with the south-eastern populations from which the modern Chinese and Taiwanese species evolved. Therefore, *F. crenata* may still be genetically closer to *F. lucida* (with possible Central Asian origins, Denk & Grimm, 2009) than *F. sylvatica* s.l. and the most southern species *F. hayatae* and *F. longipetiolata* (see Shen, 1992, for distribution maps). This contact is also reflected in the single-/low-copy genes used by Jiang et al. (2021; cf. Data S5, discussed in Section 4.6 in Data S1) but not in *LEAFY* intron data (Oh et al., 2016; Renner et al., 2016; cf. Fig. 1). Jiang et al.'s (2021) inferred ages for the diversification of Subgenus Fagus (c. 10 Ma) and speciation of *F. crenata* are, however, much too young and at odds with the rich and extensively studied fossil record of *Fagus* (cf. Denk & Grimm, 2009; Grímsson et al., 2016). First, in their dating of the multi-species coalescent (MSC), Jiang et al. (2021) assigned c. 30 and c. 20 myrs too young stem and crown ages for *Fagus*; likewise they assumed a much too young stem age for the outgroup comprising extra-tropical Castaneoideae and members of both oak subgenera (subgen. *Quercus* and *Cerris*). Jiang et al.'s (2021) inferred *Quercus* crown age is at least 25 myrs too young (cf. Hubert et al., 2014; Hipp et al., 2020). The various reticulation events between East Asian species of Subgenus Fagus indicated by Jiang et al.'s (2021) data, may lead to wrong and/or too young MSC estimates unless one infers a MSC network (see simulations in the supplementary information to Wen & Nakhleh, 2018). In addition, asymmetric introgression (a process not modelled by

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

the birth-death model), for instance via long-distance pollen dispersal (e.g. Sofiev et al., 2013), may have quickly homogenized ancestral single-/low-copy gene polymorphism.

Gömöry et al. (2018) used allozymes and approximate Bayesian computation to examine the demographic history of western Eurasian beeches and suggested that *F. sylvatica* s.str. diverged from *F. orientalis* at ~1.2 Ma in the Pleistocene (Calabrian); this scenario is supported by the leaf fossil record (e.g. Follieri, 1958; Denk et al., in press). The diversity seen in Iranian *F. orientalis* (Figs 3, 5) may well represent the original 5S-IGS diversity within the western range of the Oligocene-Miocene precursor of the *crenata-sylvatica* lineage. The poorly sorted 'Western Eurasian B' lineage reflects the common origin, the originally shared gene pool, of the western Eurasian beeches and the Japanese *F. crenata* and ongoing gene flow in the Miocene (relict 'Crenata B0'; 'Crenata B1' grade; Fig. 3). Furthermore, *F. crenata*, as the easternmost species, seems to have replaced its A-Lineage 5S-IGS variants with specific 'Crenata type B2' sequences. Originally, *Fagus* had a (near)continuous range from Japan via central Asia to Europe, and it is therefore possible that within this continuous area (extinct) *Fagus* populations acquired 5S-IGS variants that survived within the 5S gene pool of populations like those observed in Iran. Beech forests persisted throughout the entire Neogene in Iran (and in the Caucasus as well) although experiencing severe bottlenecks (Shatilova et. al., 2011; Dagtekin et al., 2020). The sharing of rare variants is consistent with this scenario as they link isolated populations and otherwise distinct species to a once common gene pool (e.g. Relict Lineage, 'European O', 'Cross-Asia' in 'Crenata B2'; Figs 3–5). These variants are not part of the terminal subtrees, but instead reflect ancient, largely lost diversity that predates the formation of the modern species and possibly even the final split between Subgenus Engleriana (represented by *F. japonica*) and Subgenus Fagus *(crenata-sylvatica* lineage). 5S-IGS data from the North American species are needed in this context: do they have high-divergent, unique variants (as in non-pseudogenic ITS copies) or do they share lineages with certain Eurasian species (as in certain single-/low-copy genes)? Denk et al. (2002) reported for instance a pseudogenous ITS clone from a Georgian (Transcaucasus) individual sharing mutation patterns otherwise exclusive and diagnostic for the North American *F. grandifolia*.

In Europe, the Pleistocene fluctuations and Holocene expansion could have triggered the secondary sorting and homogenization of both A- and B-Lineage 5S-IGS arrays: 'European A' and 'Western Eurasian B' clades include relatively high proportions of shared

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

("ambiguous") 5S-IGS variants in contrast to the highly coherent 'European B' clade (Figs 3, 4). Unhindered gene flow lasted much longer between Greek *F. orientalis* and *F. sylvatica* s.str. than between the latter and the more disjunct Iranian *F. orientalis* and/or Japanese *F. crenata*. A side effect of the higher genetic exchange between the western populations, a putatively larger active population size and more dynamic history, is the retention of ancient 5S-IGS variants that are likely relicts from a past diversification.

Future research directions: From species trees to species networks

The rapidly increasing amount of genomic data demonstrate hybridization, reticulation in general, is the rule rather than an exception in plants (Yakimowski & Rieseberg, 2014). Genomes of numerous modern angiosperm lineages have preserved signals of multiple rounds of whole genome duplications; plant genome evolution includes complex processes of chromosome fractionation and rearrangement (Nieto Feliner et al., 2020). High-Throughput Sequencing opens up new avenues for investigating ancient hybridization/reticulation and the integration of distinct parental genomic traits into progenies (e.g. Ma et al., 2019; Cruz-Nicolás et al., 2021).

The 5S intergenic spacers provide a unique source of information to investigate these multifaceted evolutionary processes (see also Grimm & Denk, 2010; Volkov et al., 2017; Garcia et al., 2020; Vozárová et al., 2021). Despite our limited sample, our HTS data captured phylogenetic signals covering 55 million years of ancient and recent reticulation forming the gene pools of modern beeches. The O-Lineage and I/X-Lineage variants found in *Fagus japonica* signify hybrid (allopolyploid) origin, and the distribution of A- and B-lineage variants in the *crenata-sylvatica* lineage reflect speciation processes predating the establishment of modern species. In stark contrast to the standard dichotomous speciation model, modelled by phylogenetic trees, beech species perpetually form and fuse. Such reticulate speciation may be the norm in wind-pollinated Fagales (*Juglans*: Zhang et al., 2019; *Quercus*: Simeone et al., 2016; McVay et al., 2017; Hipp et al., 2020; *Nothofagus*: Premoli et al., 2012; Acosta & Premoli, 2018). The shift from insect- to wind-pollination revolutionized their speciation processes, maximizing the advantages gained from hybridization and introgression. Disentangling the key aspects of (ancient) hybrids' genome architecture, discerning hybridization and introgression from incomplete lineage sorting to reconstruct the

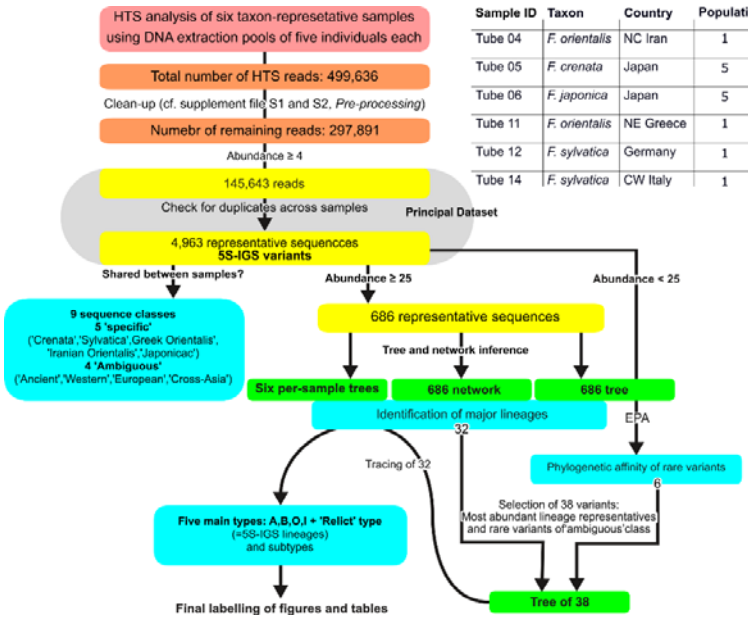
bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

multi-species coalescent network (Wen & Nakhleh, 2018), should be the ultimate goal of future research (Mallet, 2016; Martin & Jiggins, 2017).

Experimental procedure

Plant material and HTS methodology

Figure 9 shows our sampling, experimental design, and workflow. Thirty beech individuals from three species (four recognized taxa) were collected in the wild, covering 14 total populations. Four samples contained five individuals each of *Fagus sylvatica* s.l. from single populations in Germany, Italy (*F. sylvatica* s.str.), Greece, and Iran (*F. orientalis*); individual DNA extracts from five populations of *F. crenata* from Japan (Subgenus Fagus) and *F. japonica* (Subgenus Engleriana; used as outgroup) were pooled as additional samples (details provided in Data S1, S2).



Sample ID	Taxon	Country	Populations	Individuals	Site
Tube 04	*F. orientalis*	NC Iran	1	5	Noor
Tube 05	*F. crenata*	Japan	5	5	Kuromatsunai (Hokkaido), Yamagata (Honshu), Shizuoka (Honshu), Nara (Honshu), Kagoshima (Kyushu)
Tube 06	*F. japonica*	Japan	5	5	Iwate (Honshu), Gunma (Honshu), Hiroshima (Honshu), Tokushima (Shikoku), Miyazaki (Kyushu)
Tube 11	*F. orientalis*	NE Greece	1	5	Dadia
Tube 12	*F. sylvatica*	Germany	1	5	Göttingen
Tube 14	*F. sylvatica*	CW Italy	1	5	Oriolo Romano

Fig. 9. Analysis set-up and identification workflow

Species and taxa were morphologically identified (Shen, 1992; Denk 1999a, b). DNA extractions were performed from silica-gel dried leaves with the DNeasy plant minikit (QIAGEN) and quantified with a NanoDrop spectrophotometer (TermoFisher Scientific). We

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

prepared six artificial samples, consisting of pure species/taxa and/or populations, by pooling equal amounts of DNA of every individual up to a total of 20 ng per sample. Paired-end Illumina sequencing (2×300 bp) was performed by LGC Genomics GmbH using the 5S-IGS the plant specific primer pair CGTGTTTGGGCGAGAGTAGT (forward) and CTGCGGAGTTCTGATGG (reverse). Raw sequences were deposited in the Sequence Read Archive under BioProject PRJNA681175.

Bioinformatics tools and workflow

Illumina paired-end reads (raw sequence data with adapters and primers clipped) were processed using MOTHUR v.1.33.0 (Schloss et al., 2009). Contigs between read pairs were assembled using the ΔQ parameter (Kozich et al., 2013) for solving base calls differences in the overlapping region and no ambiguous bases were allowed. Reads were dereplicated and screened for chimeras using UCHIME in *de-novo* mode (Edgar et al., 2011) within MOTHUR and the unique (non-redundant) sequences (5S-IGS variants) with abundance < 4 were excluded from further analyses.

Given the absence of reference beech 5S rDNA sequences available in public repositories, the sequencing success of the target amplicon (5S-IGS) relied on BLAST searches (https://www.ncbi.nlm.nih.gov/; accessed on July 15th, 2020) within Fagales, with randomly chosen sequence reads, and reference to known literature (cf. Data S1, sections 2.2, 5).

Unique 5S-IGS sequence variants, selected based on abundance thresholds or per sample, were aligned using MAFFT v.7 (Katoh & Standley, 2013); MAFFT-generated multiple sequence alignments (MSAs) were manually checked and sequence ends trimmed using SEAVIEW v.4.0 (Gouy et al., 2010). Sequence length and GC content percentage were calculated with JEMBOSS 1.5 (Carver & Bleasby, 2003) and plotted with GGPLOT2 R package (Wickham, 2016). Occurrence and relative abundance of sequence reads in each sample were used for a preliminary identification of the obtained sequences (Data S1, section 2.3; Data S2). "Specific" classes included sequence reads (near-)exclusive (>99.95%) to one taxon/sample; sequences shared among taxonomically different samples were defined "ambiguous". A graphical visualization of the links between the 5S-IGS variants in each sample and class assignation was performed by Circos plot (Krzywinski et al., 2009).

Variants with abundance ≥ 25 (686-tip set) were used in the phylogenetic analyses (trees and networks). Maximum likelihood (ML) analyses relied on RAXML v.8.2.11 (Stamatakis,

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

2014). Trees were inferred under a GTR + Γ substitution model using the 'extended majority-rule consensus' criterion as bootstopping option (with up to 1000 bootstrap [BS] pseudoreplicates; Pattengale et al., 2009). Trees' visualization was performed in iTOL (www.itol.embl.de; Letunic & Bork, 2019) or DENDROSCOPE 3 (Huson & Scornavacca, 2013). The Neighbour-Net (NNet) algorithm implemented in SPLITSTREE4 (Bryant & Moulton, 2004; Huson & Bryant, 2006) was used to generate a planar (equal angle, parameters set to default) phylogenetic network to explore incompatible phylogenetic signals and visualize general diversification patterns (cf. Hipp et al., 2020). The NNet used simple (Hamming) uncorrected pairwise (*p*-) distances computed with SPLITSTREE4. Intra- and inter-lineage diversity, between and within main 5S-IGS types (see *Results*), was assessed with MEGA X (Kumar et al., 2018).

The phylogenetic position of variants with low abundance (<25) was evaluated with the 'Evolutionary Placement Algorithm' (EPA; Berger et al., 2011; Data S2, S3) implemented in RAxML. EPA was done by compiling MSAs for each sequence class and using the phylogenetic tree inferred from the 686-tip dataset; its outputs showed the multiple possible placement positions of each query sequence with different likelihood weights (probabilities) in all branches of the tree. Placement results (tree.jplace standard placement format) were visualized using iTOL. All phylogenetic and EPA analysis files, including used MSAs and a MSA collecting all 4693 sequences, are included in the Online Data Archive (http://dx.doi.org/10.6084/m9.figshare.16803481).

Following all phylogenetic analyses (per-sample trees, 686-tip tree and network, EPA assignation), we selected a set of 38 variants including the most abundant variants belonging to each identified lineage and the most abundant (taxon-)"specific" and "ambiguous" variants found in each sample, and additional strongly divergent and/or rare variants, for a further interpretation of the results. The autogenerated MAFFT MSA was checked and curated for all length-polymorphic regions using MESQUITE v. 2.75 (Maddison and Maddison 2011). Phylogenetic tree inference for the selected data used RAxML included or excluded two generally length-polymorphic regions (see *Results*) and 10,000 BS pseudo-replicates to establish support for competing splits. Data S4 documents the selection and includes a fully annotated, tabulated version of the 38-tip alignment.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Accession numbers

All generated raw HTS sequences are deposited in the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra) under BioProject PRJNA681175.

Conflict of interest

The Authors declare no potential sources of conflict of interest.

Acknowledgements

EDS and GWG gratefully acknowledge the support of the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig funded by the German Research Foundation (FZT 118). The research was partially supported by MIUR (Italian Ministry for Education, University and Research), Law 232/2016, "Department of excellence".

Author contribution

GWG, ACP, JRPW, MCS conceived and designed the study

SC, RP, GWG, ACP, MCS performed research

AS, E-DS, PSS, YS, NT, JRPW contributed analytical tools and materials

SC, RP, TD, GWG, ACP, MCS analysed and interpreted data

All authors supervised and wrote the manuscript

Data Accessibility

Used alignments and inference files are included in the Online Data Archive available at: http://dx.doi.org/10.6084/m9.figshare.16803481. Other relevant data are contained within the manuscript and its supporting materials. The original (reconstructed, phased haplotypes up-loaded to gene banks) and revisited data (de-constructed individual sequences) of Jiang et al. (2021) have been included in an existing Fagaceae data collection (Grimm, 2020) and can be accessed at: https://doi.org/10.6084/m9.figshare.11603547.

Supporting information

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Data S1 [PDF] Details about reasoning for using HTS data, sampling, 5S-IGS identification, pre-processing, amplicon length/GC content diversity within each sample/major type, and description of most prominent length-polymorphic patterns. Includes three supplement tables, 15 supplementary figures and two appendices. Appendix A is a summary of Data S2; appendix B includes violin and scatter plots sorted by sample and type.

Data S2 [XLSX] Basic information about samples, obtained HTS reads, and completely annotated lists for the 4,693 variants with an abundance of ≥ 4, as-is and summarised as Pivot tables. Uses auto-filter, auto-generating and auto-formatting functionality; up-to-date version of EXCEL® is recommended to properly view the file.

Data S3 [PDF] Graphical representation of the EPA analysis of 4007 5S-IGS variants with a total abundance < 25 (full results provided in Online Data Archive, folder *EPA*)

Data S4 [XLSX] Details about downstream in-depth analysis of 38 selected variants, including selection process and a fully annotated alignment in tabulated, graphically enhanced form.

Data S5 [XLSX] Tabulation and characterization of gene incongruence seen in, and phylogenetic information that can be extracted from, the 28-gene data of Jiang et al. (2021); including tree inference, bootstrapping and gene-wise model statistics.

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

References

Acosta, M. C. and Premoli, A. C. (2010) Evidence of chloroplast capture in South American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). *Molecular Phylogenetics and Evolution*, 54, 235–242. https://doi.org/10.1016/j.ympev.2009.08.008

Arkhipov, S. A., Volkova, V. S., Zolnikov, I. D., Zykina, V. S., Krukover, A. A. and Kul'kova, L. A. (2005) West Siberia. In: Velichko A. A., Nechaev V. P. (eds) Cenozoic climatic and environmental changes in Russia. *Geological Society of America Special Paper* 382, 67–88.

Babik, W., Taberlet, P., Ejsmond, M. J. A. N. and Radwan, J. (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. *Molecular Ecology Resources*, 9, 713–719. https://doi.org/10.1111/j.1755-0998.2009.02622.x

Berger, S. A., Krompass, D. and Stamatakis, A. (2011) Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. *Systematic Biology*, 60, 291–302. https://doi.org/10.1093/sysbio/syr010

Bijarpasi, M. M., Müller, M. and Gailing, O. (2020) Genetic diversity and structure of Oriental and European Beech populations from Iran and Europe. *Silvae Genetica*, 69, 55–62. https://doi.org/10.2478/sg-2020-0008

Bryant, D. and Moulton, V. (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21, 255–265. https://doi.org/10.1093/molbev/msh018

Cai, Q., Welk, E., Ji, C., Fang, W., Sabatini, F. M., Zhu, J., … Bruelheide, H. (2021) The relationship between niche breadth and range size of beech (*Fagus*) species worldwide. *Journal of Biogeography*, 48, 1240–1253. https://doi.org/10.1111/jbi.14074

Carver, T. and Bleasby, A. (2003) The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics*, 19, 1837–1843. https://doi.org/10.1093/bioinformatics/btg251

Casanovas-Vilar, I., Garcia-Porta, J., Fortuny, J., Sanisidro, Ó., Prieto, J., Querejeta, M., … Alba D. M. (2018) Oldest skeleton of a fossil flying squirrel casts new light on the phylogeny of the group. *eLife*, 7, e39270. https://doi.org/10.7554/eLife.39270

Chappell, M., Robacker, C. and Jenkins, T. M. (2008) Genetic diversity of seven deciduous azalea species (*Rhododendron* spp. section *Pentanthera*) native to the eastern United States. *Journal of the American Society for Horticultural Science*, 133, 374–382. https://doi.org/10.21273/JASHS.133.3.374

Cohen, K. M., Finney, S. C., Gibbard, P. L. and Fan, J.-X. (2013, updated) The ICS International Chronostratigraphic Chart. *Episodes*, 36, 199–204. http://www.stratigraphy.org/index.php/ics-chart-timescale

Cruz-Nicolás, J., Villarruel-Arroyo, A., Gernandt, D. S., Fonseca, R. M., Aguirre-Planter, E., Eguiarte L. E., Jaramillo-Correa, J.P. (2021) Non-adaptive evolutionary processes governed the diversification of a temperate conifer lineage after its migration into the tropics. *Molecular Phylogenetics and Evolution*, 160, 107125. https://doi.org/10.1016/j.ympev.2021.107125.

Dagtekin, D., Şahan, E. A., Denk, T., Köse, N. and Dalfes, H. N. (2020) Past, present and future distributions of Oriental Beech (*Fagus orientalis*) under climate change projections. *PLoS ONE*, 15, e0242280. https://doi.org/10.1371/journal.pone.0242280

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Darley-Hill S. and Johnson W. C. (1981) Acorn dispersal by the blue jay (*Cyanocitta cristata*). *Oecologia*, 50, 231–232. https://doi.org/10.1007/BF00348043

De Queiroz, K. (2007) Species concepts and species delimitation. *Systematic Biology*, 56, 879–886. https://doi.org/10.1080/10635150701701083

Denk, T. (1999a) The taxonomy of *Fagus* in western Eurasia, 1: *Fagus sylvatica* subsp. *orientalis* (= *F. orientalis*). *Feddes Repertorium*, 110, 177–200. https://doi.org/10.1002/fedr.19991100305

Denk, T. (1999b) The taxonomy of *Fagus* in western Eurasia, 2: *Fagus sylvatica* subsp. *sylvatica*. *Feddes Repertorium*, 110, 381–412. https://doi.org/10.1002/fedr.19991100510

Denk, T. (2003) Phylogeny of *Fagus* L. (Fagaceae) based on morphological data. *Plant Systematics and Evolution*, 240, 55–81. https://doi.org/10.1007/s00606-003-0018-x

Denk, T. (2004) Revision of *Fagus* from the Cenozoic of Europe and southwestern Asia and its phylogenetic implications. *Documenta Naturae*, 150, 1–72

Denk, T. and Grimm, G. W. (2009) The biogeographic history of beech trees. *Review of Palaeobotany and Palynology*, 158, 83–100. https://doi.org/10.1016/j.revpalbo.2009.08.007

Denk, T. and Grimm, G. W. (2010) The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. *Taxon*, 59, 351–366. https://doi.org/10.1002/tax.592002

Denk, T. and Tekleva, M. V. (2014) Pollen morphology and ultrastructure of *Quercus* with focus on Group Ilex (=*Quercus* subgenus *Heterobalanus* (Oerst.) Menitsky): implications for oak systematics and evolution. *Grana*, 53, 255–282

Denk, T., Grimm, G. W. and Hemleben, V. (2005) Patterns of molecular and morphological differentiation in *Fagus*: implications for phylogeny. *American Journal of Botany*, 92, 1006–1016. https://doi.org/10.3732/ajb.92.6.1006

Denk, T., Grimm, G., Stögerer, K., Langer, M. and Hemleben, V. (2002) The evolutionary history of *Fagus* in western Eurasia: evidence from genes, morphology and the fossil record. *Plant Systematics and Evolution*, 232, 213–236. https://doi.org/10.1007/s006060200044

Denk, T., Sami, M., Teodoridis, V., Martinetto, E. (in press) The late early Pleistocene flora of Oriolo, Faenza (Italy): assembly of the modern forest biome. *Fossil Imprint* 00: 00–00.

Edgar, R. C, Haas, B. J., Clemente, J. C., Quince, C. and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200. https://doi.org/10.1093/bioinformatics/btr381

Endress, P. K. (1977) Evolutionary trends in the Hamamelidales-Fagales-group. In: Kubitzki K. (ed.) *Flowering Plants*. Plant Systematics and Evolution/ Entwicklungsgeschichte und Systematik der Pflanzen, vol 1. Vienna, Springer, pp. 321–347.

Erichsen, E. O., Budde, K. B., Sagheb-Talebi, K., Bagnoli, F., Vendramin, G. G., Hansen, O. K. (2018) Hyrcanian forests–Stable rear-edge populations harbouring high genetic diversity of *Fraxinus excelsior*, a common European tree species. *Diversity and Distribution*, 24, 1521– 1533. https://doi.org/10.1111/ddi.12783

Fang, J. and Lechowicz, J. M. (2006) Climatic limits for the present distribution of beech (*Fagus* L.) species in the world. *Journal of Biogeography*, 33, 1804–1819. https://doi.org/10.1111/j.1365-2699.2006.01533.x

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Folk, R. A., Mandel, J. R. and Freudenstein, J. V. (2017) Ancestral gene flow and parallel organellar genome capture result in extreme phylogenomic discord in a lineage of angiosperms. *Systematic Biology*, 66, 320–337. https://doi.org/10.1093/sysbio/syw083

Folk, R. A., Soltis, P. S., Soltis, D. E. and Guralnick, R. (2018) New prospects in the detection and comparative analysis of hybridization in the Tree of Life. *American Journal of Botany*, 105, 364–375. https://doi.org/10.1002/ajb2.1018

Follieri, M. (1958) La foresta colchica fossile di Riano Romano. I. Studio dei fossili vegetali macroscopici. *Annali di Botanica*, 26, 129–142

Fradkina, A. F., Alekseev, M. N., Andreev, A. A. and Klimanov, V. A. (2005) East Siberia. In: Velichko, A. A. and Nechaev, V. P. (eds) Cenozoic Climatic and Environmental Changes in Russia. *Geological Society of America Special Paper*, 382, 105–120.

Friedman, J. and Barrett, S. C. H. (2008) A phylogenetic analysis of the evolution of wind pollination in the angiosperms. *International Journal of Plant Sciences*, 169, 49-58. https://doi.org/10.1086/523365

Friis, E. M., Crane, P. R. and Pedersen, K. R. (2011) *Early flowers and angiosperm evolution*. Cambridge, Cambridge University Press

Garcia, S., Wendel, J. F., Borowska-Zuchowska, N., Aïnouche, M., Kuderova, A. & Kovarik, A. (2020) The utility of Graph Clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids, and cryptic introgressants. *Frontiers in Plant Science*, 11, 41. https://doi.org/10.3389/fpls.2020.00041

Glenn, T. C. (2011) Field guide to next-generation DNA sequencers. *Molecular Ecology Resources*, 11, 759–769. https://doi.org/10.1111/j.1755-0998.2011.03024.x

Gömöry, D. and Paule, L. (2010) Reticulate evolution patterns in western Eurasian beeches. *Botanica Helveticae*, 120, 63–74. https://doi.org/10.1007/s00035-010-0068-y

Gömöry, D., Paule, L. and Mačejovský, V. (2018) Phylogeny of beech in western Eurasia as inferred by approximate Bayesian computation. *Acta Societatis Botanicorum Poloniae*, 87, 3582. https://doi.org/10.5586/asbp.3582

Gouy, M., Guindon, S. and Gascuel, O. (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, 27, 221–224. https://doi.org/10.1093/molbev/msp259

Grimm, G. (2020) Fagaceae collection. Dataset, *figshare*. Version 3, posted 29/9/2021. https://doi.org/10.6084/m9.figshare.11603547.v3

Grimm, G. W. and Denk, T. (2008) ITS Evolution in *Platanus* (Platanaceae): homoeologues, pseudogenes and ancient hybridization. *Annals of Botany*, 101, 403–419. https://doi.org/10.1093/aob/mcm305

Grimm, G. W. and Denk, T. (2010) The reticulate origin of modern plane trees (*Platanus*, Platanaceae) – a nuclear marker puzzle. *Taxon*, 59, 134–147. https://doi.org/10.1002/tax.591014

Grimm, G. W. and Denk, T. (2014) The Colchic region as refuge for relict tree lineages: cryptic speciation in field maples. *Turkish Journal of Botany*, 38, 1050–1066. https://doi.org/10.3906/bot-1403-87

Grimm, G. W., Denk, T. and Hemleben, V. (2007) Coding of intraspecific nucleotide polymorphisms: a tool to resolve reticulate evolutionary relationships in the ITS of beech trees (*Fagus* L., Fagaceae). *Systematics and Biodiversity*, 5, 291–309. https://doi.org/10.1017/S1477200007002459

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Grimm, G. W., Renner, S. S., Stamatakis, A. and Hemleben, V. (2007) A nuclear ribosomal DNA phylogeny of *Acer* inferred with maximum likelihood, splits graphs, and motif analysis of 606 sequences. *Evolutionary Bioinformatics*, 2, 7–22. https://doi.org/10.1177/117693430600200014

Grimm, G., Denk, T. and Hemleben, V. (2007) Evolutionary history and systematics of *Acer* section *Acer* – a case study of low-level phylogenetics. *Plant Systematics and Evolution*, 267, 215–253. https://doi.org/10.1007/s00606-007-0572-8

Grímsson, F. and Denk, T. (2007) Floristic turnover in Iceland from 15 to 6 Ma – extracting biogeographical signals from fossil floral assemblages. *Journal of Biogeography*, 34, 1490–1504. https://doi.org/10.1111/j.1365-2699.2007.01712.x

Grímsson, F., Grimm, G. W., Zetter, R. and Denk, T. (2016) Cretaceous and Paleogene Fagaceae from North America and Greenland: evidence for a Late Cretaceous split between *Fagus* and the remaining Fagaceae. *Acta Palaeobotanica*, 56, 247–255. https://doi.org/10.1515/acpa-2016-0016

Grossman, J. J. (2021) Evidence of constrained divergence and conservatism in climatic niches of the temperate maples (*Acer* L.) *Forests*, 12, 535. https://doi.org/10.3390/f12050535

Hatziskakis, S., Papageorgiou, A. C., Gailing, O. and Finkeldey, R. (2009) High chloroplast haplotype diversity in Greek populations of beech (*Fagus sylvatica* L.) *Plant Biology*, 11, 425–433. https://doi.org/10.1111／j.1438-8677.2008.00111.x

Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., … Valencia-Avalos, S. (2020) Genomic landscape of the global oak phylogeny. *New Phytologist*, 226, 1198–1212. https://doi.org/10.1111/nph.16162

Hofmann, C. C., Kodrul, T. M., Liu, X., and Jin, J. (2019) Scanning electron microscopy investigations of middle to late Eocene pollen from the Changchang Basin (Hainan Island, South China) – insights into the paleobiogeography and fossil history of *Juglans*, *Fagus*, *Lagerstroemia*, *Mortoniodendron*, *Cornus*, *Nyssa*, *Symplocos* and some Icacinaceae in SE Asia. *Review of Palaeobotany and Palynology*, 265, 41–61. https://doi.org/10.1016/j.revpalbo.2019.02.004

Hubert, F., Grimm, G. W., Jousselin, E., Berry, V., Franc, A. and Kremer, A. (2014) Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*. *Systematics and Biodiversity*, 12, 405-423. https://doi.org/10.1080/14772000.2014.941037

Huson, D. H. and Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23, 254–267. https://doi.org/10.1093/molbev/msj030

Huson, D. H. and Scornavacca, C. (2012) Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Systematic Biology*, 61, 1061–1067. https://doi.org/10.1093/sysbio/sys062

Jiang, L., Bao, Q., He, W., Fan, D.-M., Cheng, S.-M., López-Pujol, … Zhang, Z.-Y. (2021) Phylogeny and biogeography of *Fagus* (Fagaceae) based on 28 nuclear single/low-copy loci. *Journal of Systematics and Evolution.* https://doi.org/10.1111/jse.12695

Katoh, K. and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772–780. https://doi.org/10.1093/molbev/mst010

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Klein, J. T. and Kadereit, J. W. (2016) Allopatric hybrids as evidence for past range dynamics in *Sempervivum* (Crassulaceae), a western Eurasian high mountain oreophyte. *Alpine Botany*, 126, 119–133. https://doi.org/10.1007/s00035-016-0164-8

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K. and Schloss, P. D. (2013) Development of a dual index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied Environmental Microbiology*, 79, 5112–20. https://doi.org/10.1128/AEM.01043-13

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., … Marra, M. A. (2009) Circos: An information aesthetic for comparative genomics. *Genome Research*, 19, 1639–1645. https://doi.org/10.1101/gr.092759.109

Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

Lei, M., Wang, Q., Wu, Z., López□Pujol, J., Li, D. and Zhang, Z. (2012) Molecular phylogeography of *Fagus engleriana* (Fagaceae) in subtropical China: limited admixture among multiple refugia. *Tree Genetics and Genomes*, 8, 1203–1212. https://doi.org/10.1007/s11295-012-0507-6

Letunic, I. and Bork, P. (2019) Interactive Tree of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, 47, W256-W259. https://doi.org/10.1093/nar/gkz239

Li, J., Stukel, M., Bussies, P., Skinner, K., Lemmon, A. R., Lemmon, … Swenson, N. G. (2019) Maple phylogeny and biogeography inferred from phylogenomic data. *Journal of Systematics and Evolution*, 57, 594–606. https://doi.org/10.1111/jse.12535

López de Heredia, U., Mora-Márquez, F., Goicoechea, P. G., Guillardín-Calvo, L., Simeone, M. C. and Soto, Á. (2020) ddRAD sequencing-based identification of genomic boundaries and permeability in *Quercus ilex* and *Q. suber* hybrids. *Frontiers in Plant Science*, 11, 564414. https://doi.org/10.3389/fpls.2020.564414

Ma, Y., Wang, J., Hu, Q., Li, J., Sun, Y., Zhang, L., … Mao, K. (2019) Ancient introgression drives adaptation to cooler and drier mountain habitats in a cypress species complex. *Communications Biology*, 2, 213. https://doi.org/10.1038/s42003-019-0445-z

Maddison, W. P., Maddison, D. R. (2011) *Mesquite: a modular system for evolutionary analysis*. Version 2.75. https://mesquiteproject.wikispaces.com/

Mallet, J. (2013) Species, concepts of. In Levin, S. A. (ed.) Encyclopedia of Biodiversity. Volume 6. Waltham, MA, Academic Press, pp. 679–691.

Mallet, J., Besansky, N. & Hahn, M.W. (2016) How reticulated are species? *BioEssays*, 38, 140–149. https://doi.org/10.1002/bies.201500149

Manchester, S. R. and Dillhoff, R. A. (2004) *Fagus* (Fagaceae) fruits, foliage and pollen from the Middle Eocene of Pacific Northwestern North America. *Canadian Journal of Botany*, 82, 1509–1517. https://doi.org/10.1139/b04-112

Manos, P. S., Zhou, Z.-K. and Cannon C. H. (2001) Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *International Journal of Plant Sciences*, 162, 1361–1379. https://doi.org/10.1086/322949

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Marques, I., Draper, D., López-Herranz, M. L., Garnatje, T., Segarra-Moragues, J. G. and Catalán, P. (2016) Past climate changes facilitated homoploid speciation in three mountain spiny fescues (*Festuca*, Poaceae). *Scientific Reports*, 6, 36283. https://doi.org/10.1038/srep36283

Martin S. H. & Jiggins C. D. (2017) Interpreting the genomic landscape of introgression. *Current Opinion in Genetics & Development*, 47, 69–74. https://doi.org/10.1016/j.gde.2017.08.007

Mayr, G. (2005) A *Fluvioviridavis*-like bird from the Middle Eocene of Messel, Germany. *Canadian Journal of Earth Sciences*, 42, 2021–2037. https://doi.org/10.1139/e05-060

McVay, J. D., Hipp, A. L. and Manos, P. S. (2017) A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society of London Series B*, 284, 20170300. https://doi.org/10.1098/rspb.2017.0300

Meyer, H. Wand Manchester, S. R. (1997) The Oligocene Bridge Creek flora of the John Day Formation, Oregon. Oakland, CA, University of California Press.

Müller, M., Lopez, P. A., Papageorgiou, A. C., Tsiripidis, I. and Gailing, O. (2019) Indications of genetic admixture in the transition zone between *Fagus sylvatica* L. and *Fagus sylvatica* ssp. *Diversity*, 11, 90. https://doi.org/10.3390/d11060090

Nieto Feliner, G., Casacuberta, J. & Wendel, J.F. (2020) Genomics of evolutionary novelty in hybrids and polyploids. *Frontiers in Genetics*, 11, 792. https://doi.org/10.3389/fgene.2020.00792

Oh, S.-H., Youm, J.-W., Kim, Y.-I. and Kim, Y.-D. (2016) Phylogeny and evolution of endemic species on Ulleungdo Island, Korea: The case of *Fagus multinervis* (Fagaceae). *Systematic Botany*, 41, 617–625. https://doi.org/10.1600/036364416X692271

Okaura, T. and Harada, K. (2002) Phylogeographical structure revealed by chloroplast DNA variation in Japanese beech (*Fagus crenata* Blume). *Heredity*, 88, 322–329. https://doi.org/10.1038/sj.hdy.6800048

Papageorgiou, A. C., Vidalis, A., Gailing, O., Tsiripidis, I., Hatziskakis, S., Boutsios, S., … Finkeldey, R. (2008). Genetic variation of beech (*Fagus sylvatica* L.) in Rodopi (N.E. Greece). *European Journal of Forest Research*, 127, 81–88. https://doi.org/10.1007/s10342-007-0185-3

Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E. and Stamatakis, A. (2009) How many bootstrap replicates are necessary? In: Batzoglou S. (ed.) Research in Computational Molecular Biology. RECOMB 2009. *Lecture Notes in Computer Science, 5541*, 184–200. Berlin, Heidelberg, Springer. https://doi.org/10.1007/978-3-642-02008-7_13

Pavlyutkin, B. I., Chekryzhov, I. Yu. and Petrenko, T. I. (2014) *Geology and floras of lower Oligocene in the Primorye*. Vladivostok, Russian Academy of Sciences Far Eastern Branch, Far Eastern Geological Institute, Dalnauka.

Peters, R. (1997) *Beech forests*. Dordrecht, Kluwer Academic Publishers.

Piredda, R., Grimm, G. W., Schulze, E.-D., Denk, T. and Simeone, M. C. (2021) High-throughput sequencing of 5S-IGS in oaks – exploring intragenomic variation and algorithms to recognize target species in pure and mixed samples. *Molecular Ecology Resources*, 21, 495–510. https://doi.org/10.1111/1755-0998.13264

Premoli, A. C., Mathiasen, P., Acosta, M. C. & Ramos, V. A. (2012) Phylogeographically concordant chloroplast DNA divergence in sympatric *Nothofagus* s.s. How deep can it be? *New Phytologist*, 193, 261–275. https://doi.org/10.1111/j.1469-8137.2011.03861.x

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Renner, S. S., Grimm, G. W., Kapli, P. and Denk, T. (2016) Species relationships and divergence times in beeches: new insights from the inclusion of 53 young and old fossils in a birth–death clock model. *Philosophical Transactions of the Royal Society of London B*, 371, 20150135. https://doi.org/10.1098/rstb.2015.0135

Ribeiro, T., Loureiro, J., Santos, C. and Morais-Cecílio, L. (2011) Evolution of rDNA FISH patterns in the Fagaceae. *Tree Genetics and Genomes*, 7, 1113–1122. https://doi.org/10.1007/s11295-011-0399-x

Sagheb-Talebi, K., Sajedi, T. and Pourhashemi, M. (2014) *Forests of Iran: a treasure from the past, a hope for the future*. Berlin, Springer.

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., … Weber, C. F. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied Environmental Microbiology*, 75, 7537–7541. https://doi.org/10.1128/AEM.01541-09

Scornavacca, C, Delsuc, Fand Galtier, N. (2020) *Phylogenetics in the Genomic Era*. No commercial publisher, authors open access book. https://hal.inria.fr/PGE/

Shatilova, I., Mchedlishvili, N., Rukhadze, L. and Kvavadze, E. (2011) *The history of the flora and vegetation of Georgia (South Caucasus)*. Tbilisi, Georgian National Museum.

Shen, C. F. (1992) *A monograph of the genus* Fagus *Tourn. ex L. (Fagaceae)*. PhD Thesis, The City University of New York, USA.

Shrestha, N., Wang, Z., Su, X., Xu, X., Lyu, L., Liu, Y., … Feng, X. (2018) Global patterns of *Rhododendron* diversity: the role of evolutionary time and diversification rates. *Global Ecology and Biogeography*, 27, 913–924. https://doi.org/10.1111/geb.12750

Simeone, M. C., Cardoni, S., Piredda, R., Imperatori, F., Avishai, M., Grimm, G. W. and Denk, T. (2018) Comparative systematics and phylogeography of *Quercus* Section *Cerris* in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ*, 6, e5793. https://doi.org/10.7717/peerj.5793

Simeone, M. C., Grimm, G. W., Papini, A., Vessella, F., Cardoni, S, Tordoni, E., … Denk, T. (2016) Plastome data reveal multiple geographic origins of *Quercus* Group Ilex. *PeerJ*, 4, e1897. https://doi.org/10.7717/peerj.1897

Sofiev, M., Siljamo, P., Ranta, H., Linkosalo, T., Jaeger, S., Rasmussen, A., … Kukkonen, J. (2013) A numerical model of birch pollen emission and dispersion in the atmosphere. Description of the emission module. *International Journal of Biometeorology*, 57, 45–58. https://doi.org/10.1007/s00484-012-0532-z

Stamatakis, A. (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30, 1312-1313. https://doi.org/10.1093/bioinformatics/btu033

Steppan S. J., Storz, B. L. and Hoffmann, R. S. (2004) Nuclear DNA phylogeny of the squirrels (Mammalia: Rodentia) and the evolution of arboreality from c-myc and RAG1. *Molecular Phylogenetics and Evolution*, 30, 703–719. https://doi.org/10.1016/S1055-7903(03)00204-5

Suarez□Gonzalez, A., Hefer, C.A., Christe, C., Corea, O., Lexer, C., Cronk, Q.C.B. and Douglas, C.J. (2016) Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, 25, 2427–2442. https://doi.org/10.1111/mec.13539

bioRxiv preprint doi: https://doi.org/10.1101/2021.02.26.433057; this version posted October 19, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC 4.0 International license.

Symonová, R. (2019) Integrative rDNAomics – Importance of the oldest repetitive fraction of the eukaryote genome. *Genes*, 10, 345. https://doi.org/10.3390/genes10050345

Volkov, R. A., Panchuk, I. I., Borisjuk, N. V., Hosiawa-Baranska, M., Maluszynska, J. & Hemleben. V. (2017) Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biology* 17, 21. https://doi.org/10.1186/s12870-017-0978-6

Vozárová, R., Herklotz, V., Kovarík, A., Tynkevich, Y. O., Volkov, R. A., Ritz, C. M. & Lunerová J. (2021) Ancient origin of two 5S rDNA families dominating in the Genus Rosa and their behavior in the Canina-Type meiosis. *Frontiers in Plant Sciences*, 12, 643548. https://doi.org/10.3389/fpls.2021.643548

Wen, D. and Nakhleh, L. (2018) Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Systematic Biology*, 67, 439–457. https://doi.org/10.1093/sysbio/syx085

Wessinger, C. A. (2021) From pollen dispersal to plant diversification: genetic consequences of pollination mode. *New Phytologist*, 229, 3125–3132. https://doi.org/10.1111/nph.17073

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. New York, Springer.

Willyard, A., Cronn, R. and Liston, A. (2009) Reticulate evolution and incomplete lineage sorting among the Ponderosa Pines. *Molecular Phylogenetics and Evolution*, 52, 498–511. https://doi.org/10.1016/j.ympev.2009.02.011

Worth, J. R. P, Ujino, T. I., Grimm, G. W., Wei, F.-J., Simeone, M. C., Li, P., …, Tomaru, N. (2021) Chloroplast genome sequencing reveals complex patterns of ancient and recent chloroplast sharing in Japanese *Fagus*. Presentation, *ResearchGate*, https://doi.org/10.13140/RG.2.2.10679.24485

Yakimowski, S. B. and Rieseberg, L. H. (2014) The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. *American Journal of Botany*, 101, 1247-1258. https://doi.org/10.3732/ajb.1400201

Yan, M., Liu, R., Li, Y., Hipp, A. L., Deng, M., Xiong, Y. (2019) Ancient events and climate adaptive capacity shaped distinct chloroplast genetic structure in the oak lineages. *BMC Evolutionary Biology*, 19, 202. https://doi.org/10.1186/s12862-019-1523-z

Yang, J., Takayama, K., Youn, J. S., Pak, J. H. and Kim, S. C. (2020) Plastome characterization and phylogenomics of East Asian beeches with a special emphasis on *Fagus multinervis* on Ulleung Island, Korea. *Genes*, 11, 1338. https://doi.org/10.3390/genes11111338

Zhang, B.-B., Xu, L.-L., Li, N., Yan, P.-C. Jiang, X. H., Woeste, K. E., …, Bai W.-N. (2019) Phylogenomics reveals an ancient hybrid origin of the Persian Walnut. *Molecular Biology and Evolution* 36, 2451–2461. https://doi.org/10.1093/molbev/msz112

Zhang, Z.-Y., Wu, R., Wang, Q., Zhang, Z.-R., Lopez-Pujol, J., Fan, D.-M. and Li, D. Z. (2013) Comparative phylogeography of two sympatric beeches in subtropical China: species specific geographic mosaic of lineages. *Ecology and Evolution*, 3, 4461–4472. https://doi.org/10.1002/ece3.829