High-Throughput Sequencing of 5S-IGS rDNA in *Fagus* L. (Fagaceae) reveals complex evolutionary patterns and hybrid origin of modern species

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; mcsimeone@unitus.it

Abstract

Standard models of speciation assume strictly dichotomous genealogies in which a species, the ancestor, is replaced by two offspring species. The reality is more complex: plant species can evolve from other species via 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. Here, we provide the first high-throughput sequencing (HTS) 5S intergenic spacer (5S-IGS) data 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 long history of repeated incomplete lineage sorting and lineage mixing since the early Cenozoic of two or more paralogous-homoeologous 5S rDNA lineages. Extant species of *Fagus* are genetic mosaics and, at least to some part, of hybrid origin.

Keywords

Fagus; 5S rDNA; High-Throughput Sequencing; Hybrid evolution; Genetic diversity

Significance statement

The evolution of extra-tropical tree genera involves dynamic speciation processes. High-throughput sequencing data of the multi-copy, potentially multi-locus 5S rDNA reveal 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 35 million years of ongoing reticulate evolution.

Introduction

Fagus L. (Fagaceae) 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 is monophyletic and might have originated at high latitudes during the Paleocene (western Greenland, northeast Asia; Fradkina et al., 2005; Denk & Grimm, 2009; Grímsson et al., 2016). It is currently subdivided into two informal subgenera corresponding to reciprocally monophyletic lineages (Shen, 1992; Denk et al., 2005). "Subgenus Engleriana" currently 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. 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, as inferred from a fossilized birth–death (FBD) model with an extensive set of 53 fossils (Renner et al., 2016), including the oldest unambiguous macrofossil record of the genus (Manchester & Dillhoff, 2004). While the lineage leading to the modern genus is at least 82–81 myrs old (Grímsson et al., 2016), the 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). These dynamic migration and speciation histories left multifarious morphological and molecular imprints on modern members of the genus.

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, 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 the entire genus (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). This is not surprising; ancient and reiterated hybridization

events can be assumed for the evolution of *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). Accordingly, plastid DNA variation has shown to be not sorted by speciation but rather by geography in all broadly sampled groups of Fagales studied so far (Fagaceae: Simeone et al., 2016; Yan et al., 2019; Nothofagaceae: Acosta & Premoli, 2010; Premoli et al., 2012).

Nuclear ribosomal DNA spacers, organized in multi-copy arrays, have an unsurpassed potential to detect past and recent reticulation events (e.g. Grimm & Denk, 2008, 2010; Pilotti et al., 2009; de Castro et al., 2013 for *Platanus*). In two subfamilies of Fagaceae, Quercoideae (oaks, *Quercus*) and Fagoideae (*Fagus*), the most widely used nuclear marker for intra-generic studies in plants, the ITS region of the 35S rDNA cistron, has shown only limited discrimination capacity while requiring extensive cloning because of substantial intra-genomic, length- and sequence-polymorphism (Denk et al., 2002, 2005; Denk & Grimm, 2010). Cloning and special methodological frameworks to extract useful phylogenetic signals and infer evolutionary patterns (e.g. Göker & Grimm, 2008; Potts et al., 2014) are also required for the shorter but typically more divergent and phylogenetically informative non-transcribed intergenic spacers of the 5S rDNA (see for instance Forest et al., 2005; Denk & Grimm, 2010; Simeone et al., 2018). This marker has never been used for large-scale genetic studies in *Fagus* but it is known to consist of two paralogous loci in *F. sylvatica* (Ribeiro et al., 2011) which may lead to substantial intra-genomic variation.

The increasing availability of inexpensive High-Throughput Sequencing (HTS) approaches is now boosting new efforts into research questions that were previously considered too expensive, labour-intensive, and/or inefficient (Babik et al., 2009; Glenn, 2011). Amplicon sequencing of loci with high information content (target sequencing) substantially reduces costs, as it makes cloning unnecessary and many individuals can be combined in the same sequencing run (Ekblom & Galindo, 2010). In a recent study, Piredda et al. (2021) analysed 5S-IGS HTS amplicon data generated for *Quercus* (Fagaceae) and demonstrated the great potential of this technique for inspecting range-wide diversity patterns and for assessing inter- and intra-species relationships. 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, climaxing in cooler versions of *Cfb* climates (i.e. warm temperate, fully humid climates with warm summer; Peters, 1997; Grimm & Denk, 2012; Kottek et al., 2016; Peel et al., 2017). 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 likely involved complex reticulations, facilitating or obscuring speciation processes during the different phases of beech evolution (Gömöry et al., 2007, 2018; Gömöry & Paule, 2010; Müller et al., 2019).

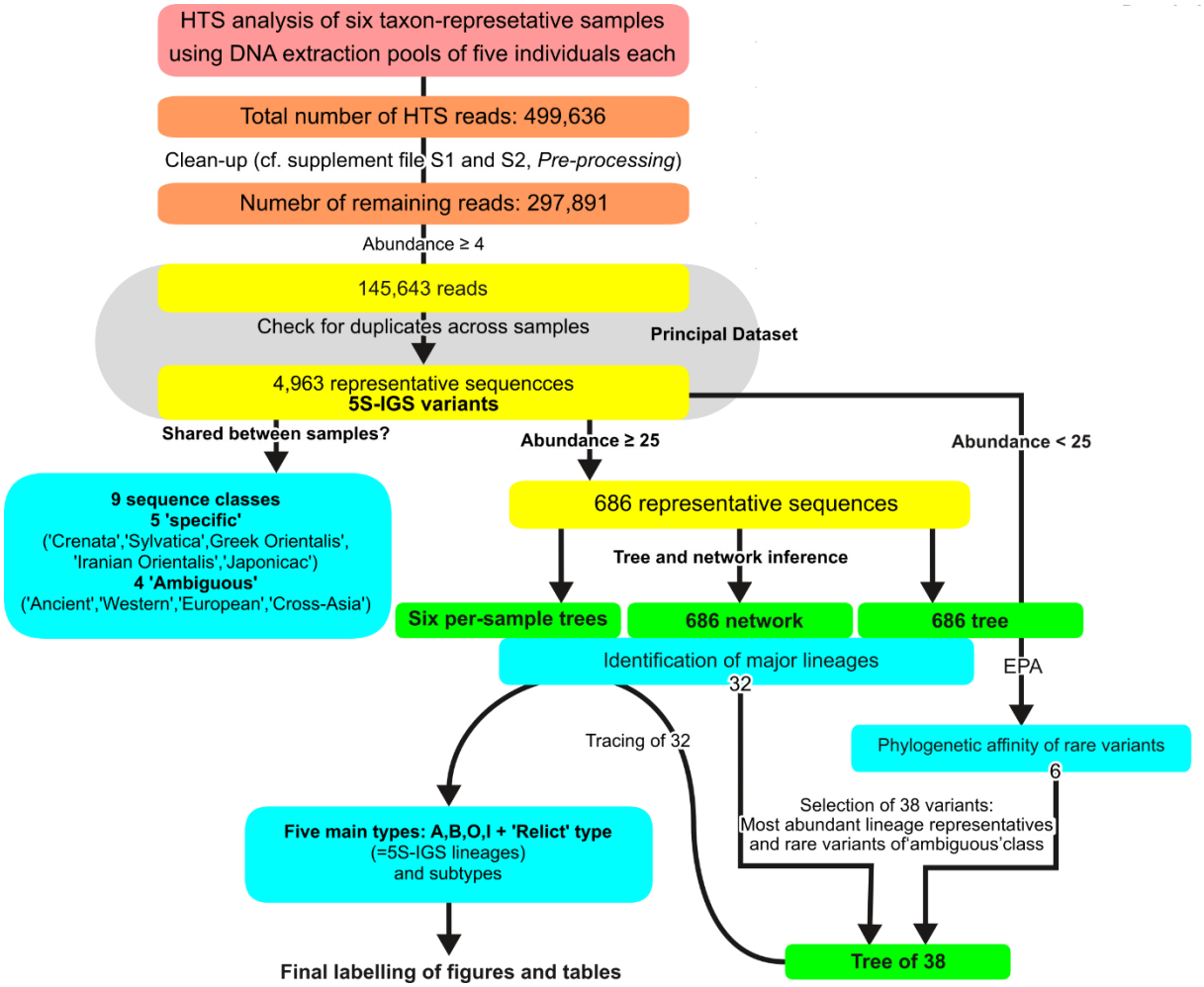
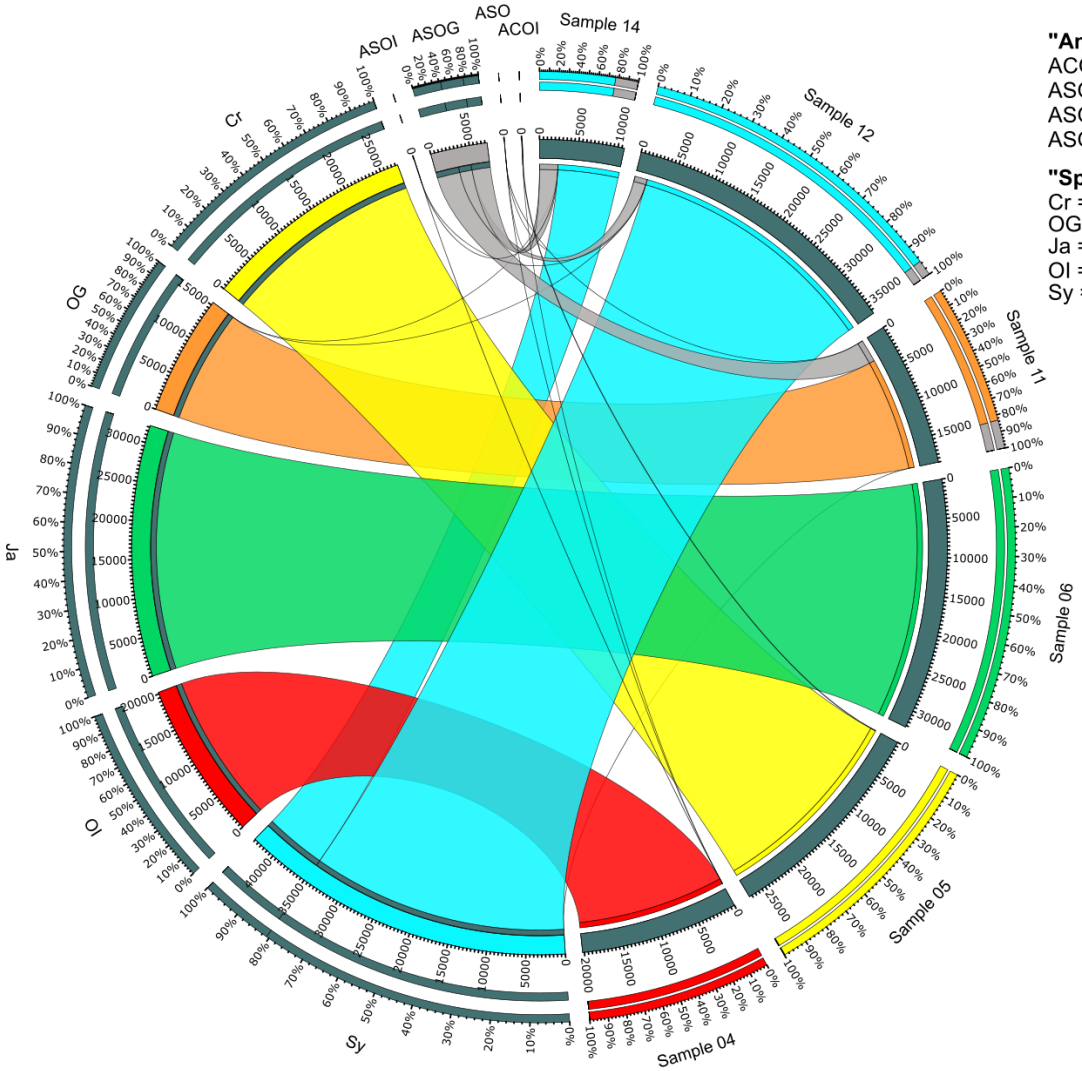


Figure 1 | Analysis set-up and identification workflow (cf. *Experimental Procedure*)



"Ambiguous" sequence classes
ACOI = "Cross-Asia", shared by *F. crenata* and Iranian *F. orientalis*
ASO = "Western", shared across *F. sylvatica* s.l.
ASOG = "European", found only in *F. sylvatica* s.str. and Greek *F. orientalis*
ASOI = "Ancient", shared by disjunct *F. sylvatica* s.str. and Iranian *F. orientalis*

"Specific" sequence classes
Cr = exclusive to *F. crenata*
OG = exclusive to Greek *F. orientalis*
Ja = exclusive to *F. japonica*
OI = exclusive to Iranian *F. orientalis*
Sy = exclusive to *F. sylvatica* s.str

Figure 2 | Map of 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 5S-IGS variants exclusively found in a single taxon sample (or two samples, in the case of *F. sylvatica* s.str.)

Results

Reads pre-processing and removal of unique variants with abundance < 4 yielded 145,643 sequences, corresponding to 4,693 representative 5S-IGS variants (Fig. 1; work details, statistics and sequence structural features reported in Cardoni et al., 2021, File S1, S2). Of these, 686 had an abundance of ≥ 25 and were included in the phylogenetic analyses; this left 4007 variants to be placed using EPA (File S2, S3). Analyses of 38 selected sequences further clarified the evolutionary relationships among the detected 5S-IGS lineages (File S4; cf. Cardoni et al. 2021).

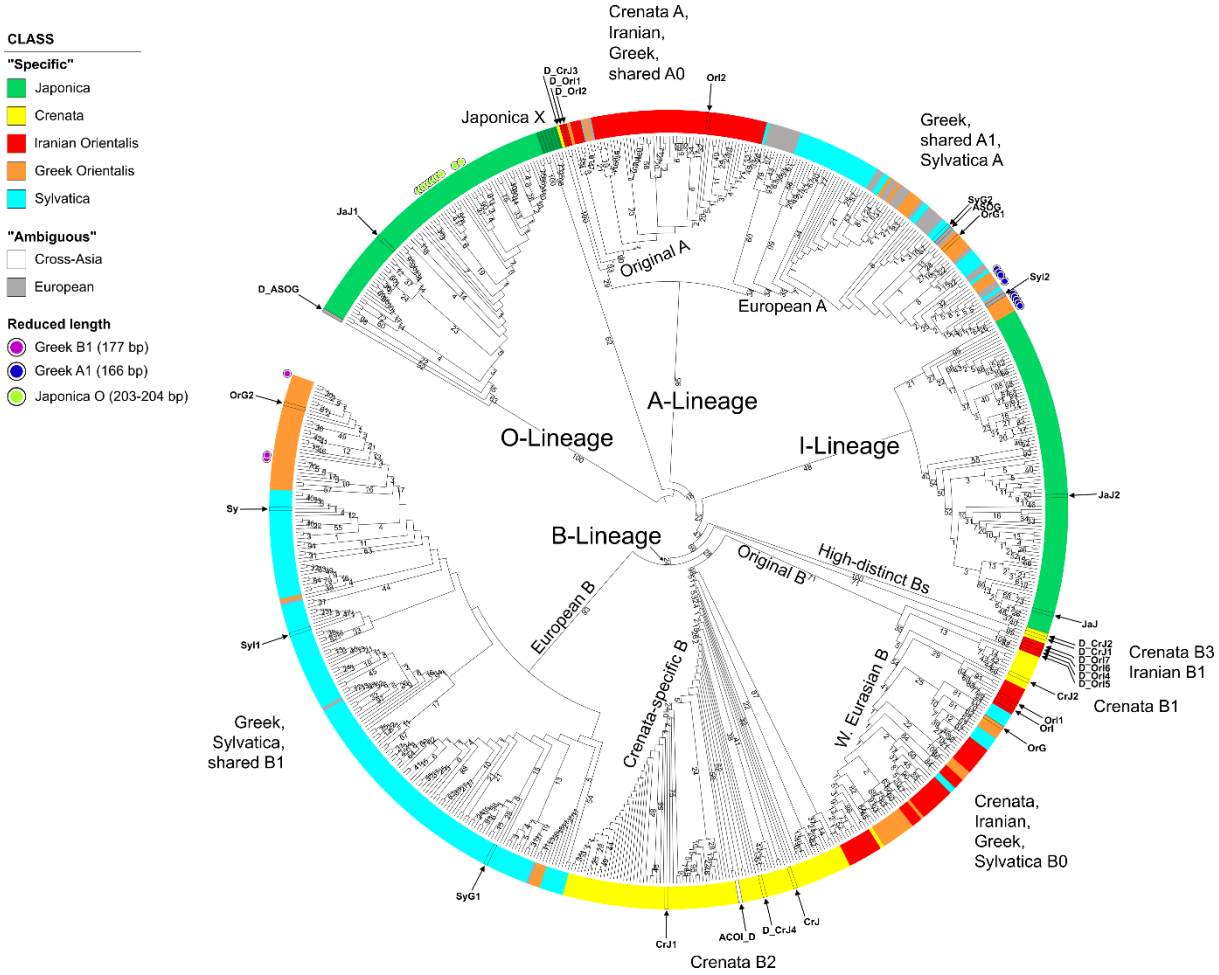


Figure 3 | Circular cladogram; based on maximum likelihood (ML) tree inference and bootstrap (BS) analysis (550 BS pseudoreplicates) of the 686-sequence matrix including all 5S-IGS variants with a total abundance ≥ 25. The tree is rooted on the genetically most distinct O-type lineage; numbers at branches 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.

686-tip 5S-IGS backbone phylogeny

We first categorized the obtained 4,693 5S-IGS sequences based 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.) of the samples representing the same taxon: "japonica", "crenata", "Iranian orientalis", "Greek orientalis", and "sylvatica". In addition, we identified four "ambiguous" classes, i.e. sequences shared among different species or taxa. Five main 5S-IGS lineages were defined based on the phylogenetic analysis of 686 sequences with abundance ≥ 25 (Fig. 3).

The two ingroup 5S-IGS main types, labelled A- and B-type, form distinct clades (BS = 95/47); they were established prior to speciation within the *F. crenata – F. sylvatica* s.l. lineage (in short: *crenata-sylvatica* lineage). The ML tree and NNet network (Fig. 4) highlight the bimodality of the five "specific" sequence classes ('japonica', 'crenata', 'sylvatica', 'Greek orientalis' and 'Iranian orientalis'). Most shared variants classified as "ambiguous" are part of the 'European A'-type lineage (BS = 34); while Iranian A-type 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-types (Fig. 4; Table 1), which comprise two genetically coherent clades ('European B' and 'Crenata B2') in addition to a poorly sorted, ambiguously supported clade ('Original B'). The ancestor of the *crenata-sylvatica* lineage must have been polymorphic and the modern pattern the result of 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 | Estimates of mean inter- and intra-lineage 5S-IGS divergence (standard error shown above the diagonal). N = sample size (number of non-identical 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	

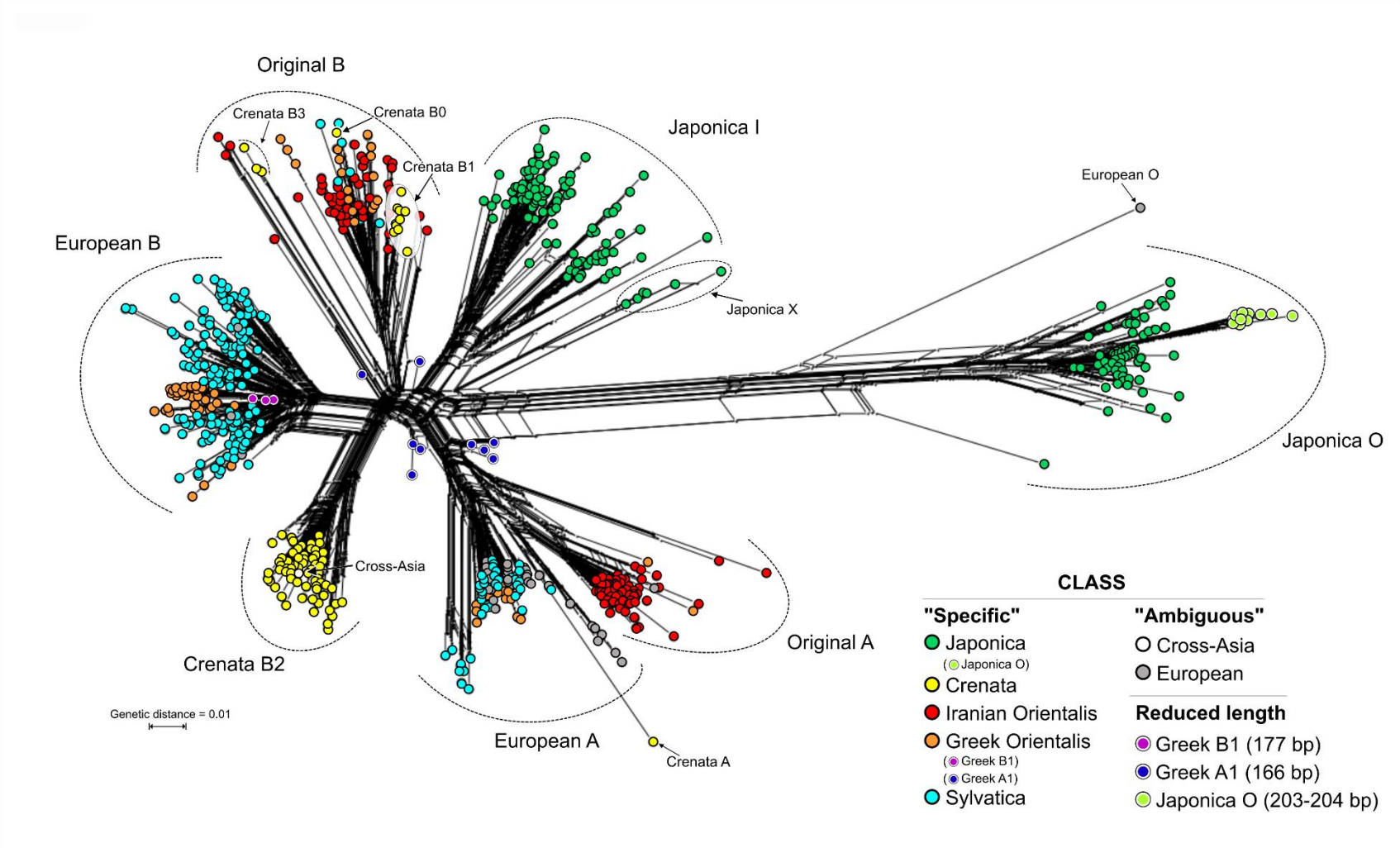


Figure 4 | Neighbour-net for the 686-sequence matrix, inferred from uncorrected (Hamming) pairwise distances. Neighbourhoods defined by well-defined interior "trunks" relate to prominent sorting events (bottleneck situations; evolutionary jumps) leading to coherent 5S-IGS lineages with high (near-unambiguous)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

There is no evidence for ongoing lineage mixing in Japan between *F. japonica* ('subgenus Engleriana') and *F. crenata* ('subgenus Fagus'). One of the *F. japonica* types, 'Japonica I', is substantially more similar to both ingroup main types (A- and B-type) than the other dominant type ('Japonica O'; Fig. 4). One "ambiguous" variant shared by western Eurasian beeches ('European O') is part of the O-type lineage; a small clade of *F. japonica*-exclusive variants ('Japonica X') nests between the A-B-I clade and the O-clade.

Fagus crenata shares a relative recent common origin with the western Eurasian beeches, represented by the 'Original B'-type (BS = 71) with two subgroups. The '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) have persisted in Italian, Greek and Iranian populations. In contrast to its western relatives, *Fagus crenata* lacks non-degraded type A variants (Table 2; File 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' 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'; Fig. 2). One sequence ('Crenata A') is placed within the 'Original A' clade, as sister to all other variants (Fig. 3), which may be a tree-branching artefact (Fig. 4).

The only other subtype showing the same level of genetic coherence (Fig.4), is the 'European B'-type (BS = 93), shared exclusively by *F. sylvatica* s.str. and Greek *F. orientalis* (Table 2). Given its distinctness, the 'European type B' likely reflects 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 the continuum, Iranian *F. orientalis* appears to be most isolated and ancient/ancestral with respect to *F. crenata* and *F. japonica.*

Table 2 | Main 5S-IGS sequence types observed in our sample and their evolutionary interpretatation

Type	Taxonomic coverage	Evolutionary interpretation
O	*F. japonica,* co-dominant, mostly inconspicuous *Crenata-sylvatica* lineage, very rare, degraded	Evidences ancient hybridisation/polyploidisation 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, mostly degraded	Shared ancestral type, most diverse in Iranian *orientalis*
B	*Crenata-sylvatica* lineage, frequent, mostly inconspicuous	Lineage-specific types, diverse and ± sorted
Original B[a]	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 ancient 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; ± degraded	Reminiscent of ancient speciation processes or reticulation
X	*F. japonica,* rare, inconspicuous	Unclear; potentially linked to non-sampled Chinese spp. of 'Subgenus Fagus'

[a] Two subtypes: Crenata B1, Western Eurasian B (incl. Crenata B0)

Very short 'Orientalis' variants are deeply embedded within the 'European A' and 'European B' clades, while short 'japonica' variants are all but one O-type (File S1, section 4.1). The distance-based NNet placed all short variants next to the centre of the graph. 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 ('Japonica I', A- and B-types).

Framework phylogeny (Fig. 5)

The ML trees for the 38 selected sequences (representing the most abundant variants within each sample, strongly deviating variants, and variants shared across species) highlight the deep split between outgroup (type O) and ingroup 5S-IGS variants (I-, A- and B-types). The mid-point root corresponds to the split between type O and the A-B-I clade. Tip-pruning as well as the elimination of the indiscriminative 'Japonica X' lineage led to a substantial increase in backbone branch support: the divergence between *F. japonica* ingroup variants (type I) and

B-type(s) is placed after the isolation of the A-type lineage (BS = 99/84). A group of sequentially unique, rare shared variants showing few to substantial signs of sequence degradation forms a distinct clade. This "relict lineage" is placed between the outgroup subtree (comprising the 'Japonica O'-type and sequence-degraded "ambiguous" western Eurasian variants) and the A-B-I subtree.

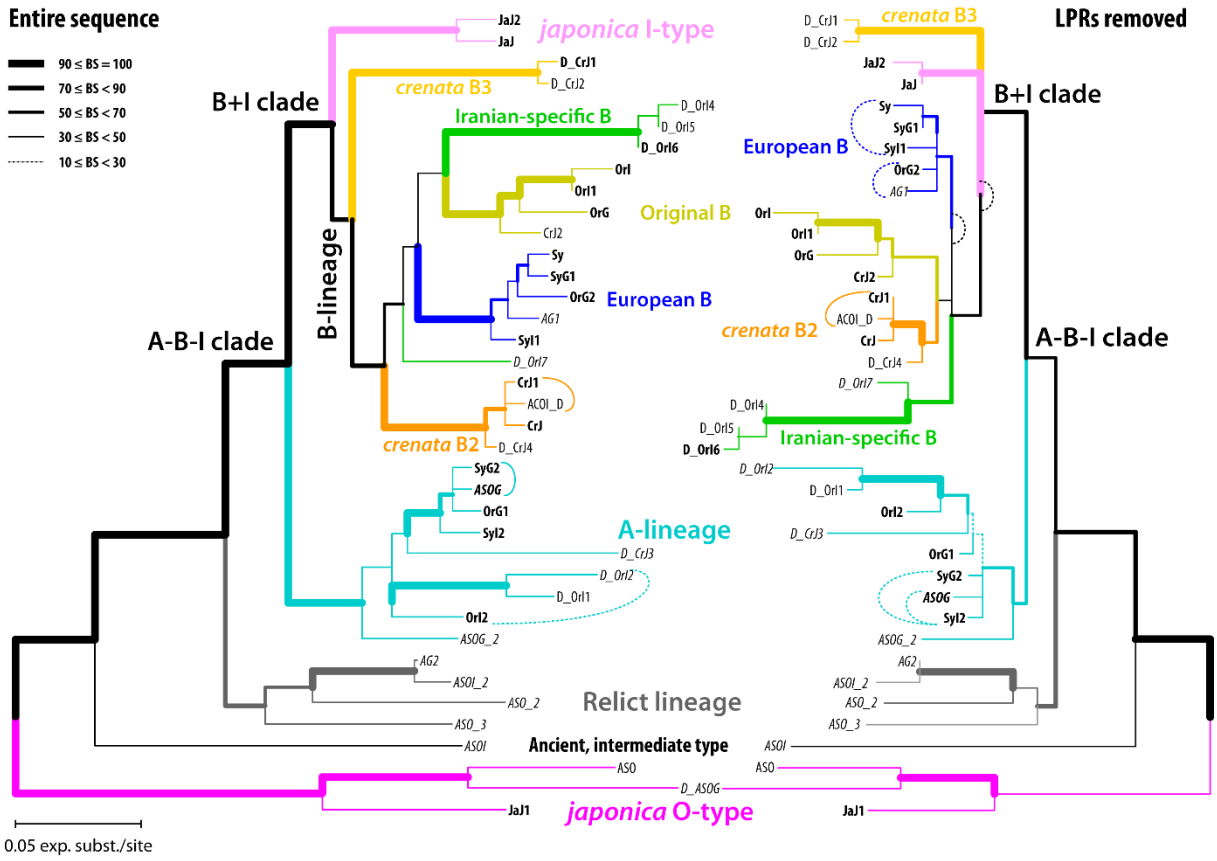


Figure 5 | Maximum likelihood phylograms inferred from the selected 38-sequence matrix; including the most common variants of each main lineage, and rare, shared variants (labelled as "*A*…") as well as high-divergent variants ("*D_*…"). Left tree was inferred with generally length-polymorphic regions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

In-depth sequence structure analysis shows that this placement is only partly due to potential ingroup-outgroup long-branch attraction. While both subtrees (outgroup and 'relict' type) include variants most different from the ingroup consensus, and branch support is higher when length-polymorphic regions are included, the rare shared types also have an increased number of mutational patterns, which appear to be primitive within the entire A-B-I lineage. Thus, they may represent relict variants; ancestral copies still found in 5S rDNA arrays that have been subsequently replaced and eliminated within the *crenata-sylvatica* lineage (Table 2). Another

important observation is that the long-branching A- and B-type variants found exclusively in the Iranian sample constitute strongly divergent, genetically coherent lineages as well as two of the low-abundant, high-divergent *F. crenata*-unique variants. While some rare, shared ingroup variants ('relict lineage', 'Crenata A', some 'European O') show clear signs of sequence degradation, others are inconspicuous (no pseudogenous mutations in flanking 5S rDNA) and can be highly similar to the most abundant variants.

Bimodality and differences between 5S-IGS populations in each sample (Fig. 6)

The unrooted, single-sample ML trees resolve two 5S rDNA clusters (5S main types) with high bootstrap values (71–100) in all studied samples. The main splits reflect the two most common main types: I- and O-type in the outgroup *F. japonica*; and A- and B-type 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-type 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 each population). These intermediate sequences do not show any structural peculiarity, except for the 'Crenata A'-type variant showing a reduced GC content (34.3%; File S1, section 4.2 and appendix B). The outgroup-type 'European O' variants represent the longest branched sequences in the Greek *F. orientalis* (Greece) and both *F. sylvatica* s.str. samples. As a trend, the variants within the B-type and I-type 5S rDNA subtrees show a higher divergence and phylogenetic structure than found within the A- and O-type subtrees. For instance, in Greek *F. orientalis* and *F. sylvatica* s.str., the 'B type' subtree includes a distinct, highly supported clade.

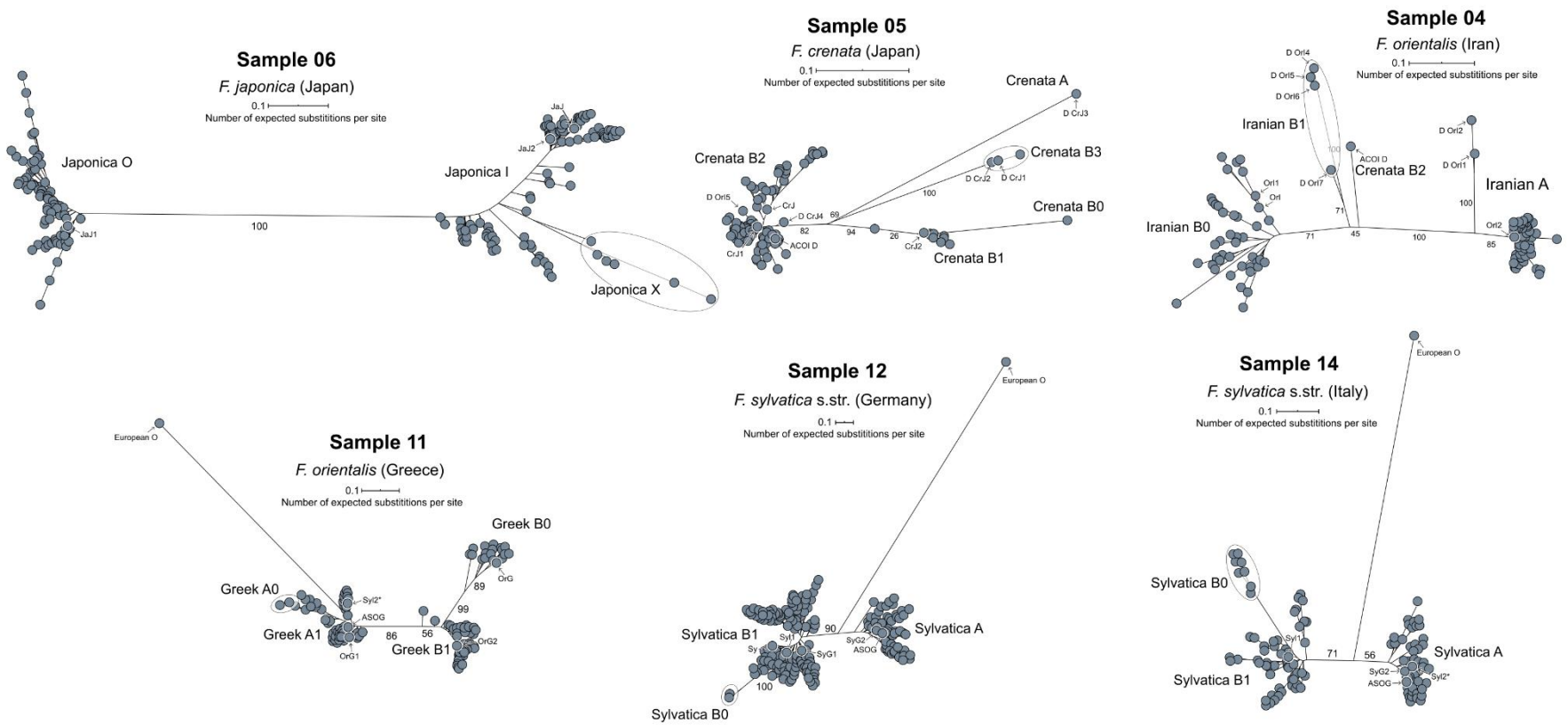


Figure 6 | Sample-wise maximum likelihood (ML) trees; illustrating the bimodality of the 5S-IGS pool in each sample. Numbers give ML bootstrap support (based on 1000 pseudoreplicates) for selected major phylogenetic splits. Subtree and individual labels refer to lineages and tips introduced in Figs 3–5.

Using EPA (Evolutionary Placement Algorithm), we assessed the phylogenetic affinity of all variants with an abundance ≥ 4 not included in the 686-tip matrix (File S2, S3). In general, 5S-IGS arrays have a high capacity to conserve signal from past reticulations and deep divergences: types placed by EPA on the branch of the 'European O'-type variant, representing a distinct but degrading and rare sister lineage of the *F. japonica* type O (cf. Table 2; Figs 3, 4), can be found in all samples of the *crenata-sylvatica* lineage (File S1, appendix A). In contrast, Type X variants are exclusive to *F. japonica*. The single, low-abudant type I variant 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 lineage. The 'relict lineage', phylogenetically in-between O- and A-/B-/I-types (cf. Fig. 5), is represented in all samples of the *crenata-sylvatica* lineage as well. Despite showing signs of sequence degradation in the flanking 5S rRNA gene regions, its GC-contents (35.1–40.0%) largely range between the median values of type O and type B/I. Hence, they match the range of A-types, and are also of the same length than most A- and B-types (File S1, appendix B).

Compilation of length diversity and GC content per main (most frequent) type and sample (Fig. 7) shows that the generally longer A-types have lower GC content than the B-type in each species; the I-B clade corresponds to a similar GC content and sequence length in the *F. japonica* type I and *crenata-sylvatica* lineage type B. Highest GC contents are found in *F. japonica* type O, which also represents the longest 5S-IGS type. While GC-richer B-type variants are more frequent, the GC-poorer A-type variants make up a higher portion of the HTS reads corresponding to rare sequence variants. In *F. crenata*, A-type 5S-IGS variants are nearly absent while in Iranian *F. orientalis* they show the highest diversity: 386 species-unique variants (40% of all A-type variants) with a total abundance of 8466 (34%). The opposite can be observed for the type B variants: while they are ± equally abundant in the Iranian *F. orientalis* sample, they are 1.5 to 2-times more frequent in Greek *F. orientalis* and Italian *F. sylvatica* samples, and approach the largest majority in the German *F. sylvatica* sample. Apparently, the GC-richer B-types subsequently replaced A-types in the genomes of the *crenata-sylvatica* lineage, while in *F. japonica* their sister lineage, type I variants, outnumbered the GC-richer, sequentially more complex O-types.

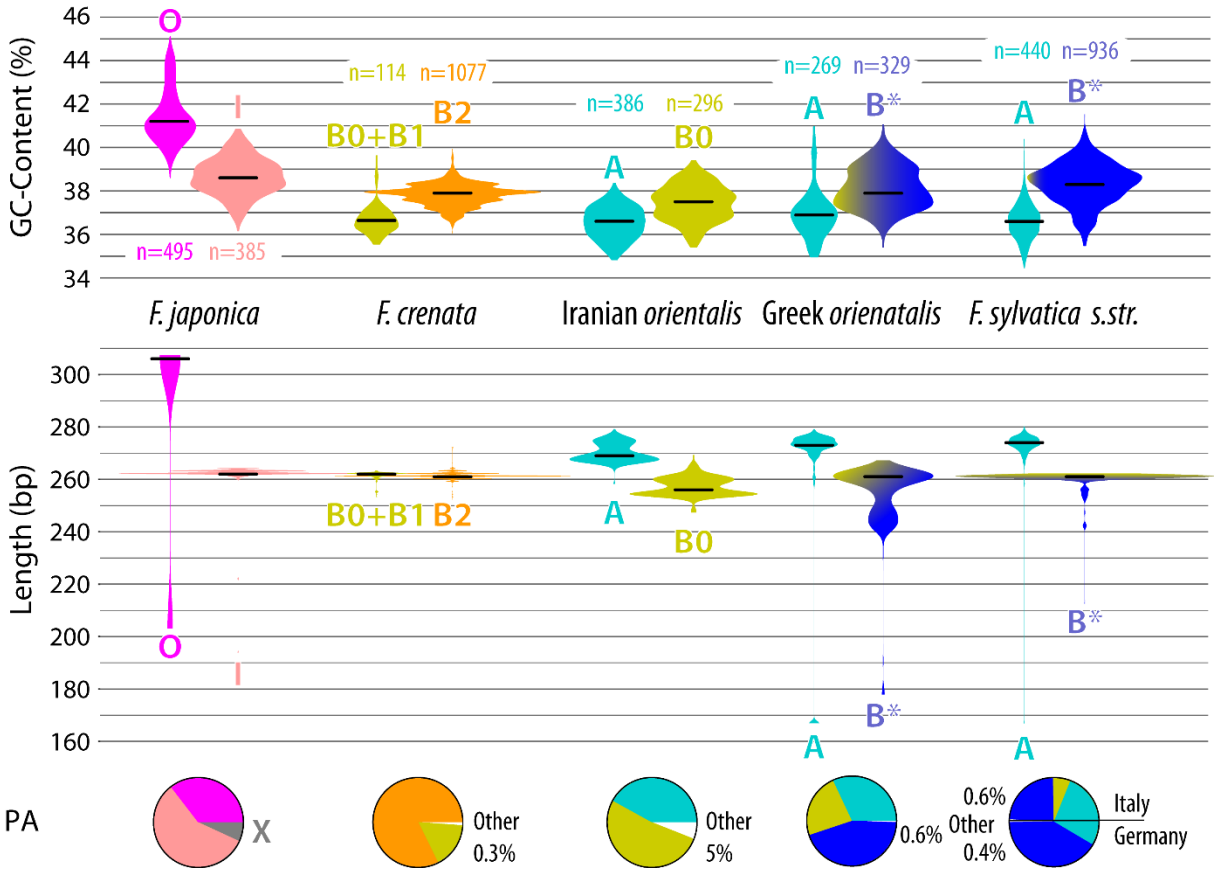


Figure 7 | Amplicon GC content and length violin plots for the (co-)dominant lineages/main 5S-IGS types found in each sample. Horizontal black lines give the median value for each main type (based on all variants of the respective lineage). Width of violin plots adjusted to visualise the relative proportion (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 File 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).

Discussion

5S-IGS divergence and paralogy/homeology

With the exception of few variants ('Crenata A', relict lineage variants, partially deleted very short O-types), there is no evidence for sequential decay. The relatively low GC contents (range: 33.2 – 45.1%) as compared to other Fagaceae such as oaks, appear to be characteristic of nuclear spacers of *Fagus*, and its nucleome in general (NCBI GenBank acc. no. QCXR00000000.1 and BKZX00000000.1; not yet annotated, accessed on 25/11/2020). Thus, pseudogeny can be largely ruled out as the cause for the detected 5S-IGS divergence. Rather, the presence of two sTable co-dominant main sequence clusters in each sample (Fig. 6; Tables 2, 3) fits with the only available cytogenetic data for *Fagus* (Ribeiro et al., 2011), showing two paralogous, potentially functional 5S rDNA pericentromeric loci (and four terminal 35S rDNA loci) in *F. sylvatica*. This is a rare feature within Fagaceae, usually showing only one 5S rDNA locus (www.plantrdnadatabase.com; accessed 15/08/2020, Chokchaichamnankit & Anamthawat-Jonsson, 2015). Additional or odd numbers of (unpaired) loci were found in single individuals and attributed to hybridisation and auto-polyploidisation (Chokchaichamnankit et al., 2008, Ribeiro et al., 2011). Comprehensive data for comparisons with other families in the Fagales are currently lacking, with the exception of *Corylus* (Betulaceae) showing a single 5S locus and much lower intra-individual, intra-specific divergence (Forest & Bruneau, 2000).

In *F. japonica*, the longer, GC-rich O-type and shorter, GC-normal I-type, represent two highly divergent, only distantly related lineages; their divergence (Table 1) parallels that between subgenera of oaks (Denk & Grimm, 2010; Simeone et al., 2018; Piredda et al., 2021 for first HTS data) and between genera in Betulaceae (Forest et al., 2005). Both types are abundant: although HTS results cannot be generally considered quantitative (Lamb et al., 2019), I- and O-type variants appear co-dominant, while type X variants, representing a second ingroup-related lineage of ambiguous phylogenetic affinity, are rarer (Fig. 7; File S1, section 4.3). This situation is similar to the one reported for cloned ITS data: 35S rDNA-cistron ITS regions of *F. japonica* and its sister species *F. engleriana* (mainland China) and *F. multinervis* (Ullung Do, South Korea) show extreme length and sequence heterogeneity with up to three divergent main sequence types, while the ITS of the *F. crenata-sylvatica* lineage is more homogeneous and poorly sorted (Denk et al., 2005; Grimm et al., 2007).

In the 5S-IGS of the *crenata-sylvatica* lineage, the length and sequence differences are less pronounced, with the longer A-type variants being generally less derived and less abundant than the shorter, much more abundant and more diverse B-type variants (Figs 3–5; File S1, appendix A). The divergence between these two types is also comparable to intra-sectional diversity in oaks and genus-level diversity in Betulaceae. The *crenata-sylvatica* samples exhibit increasing dominance of type B over type A (Fig. 7): weakly developed in Iranian *F. orientalis*, increased in European samples following a (south)east/(north)west gradient (Greece – Italy – Germany), and strongest in *F. crenata* (type A almost absent). In addition, putative relicts (Fig. 5) can be found that lack/mix features of A- and B-type variants (File S4) or are clearly related to the outgroup variants, type O ('European O'; Figs 3, 4, 6). Hence, we detected up to three major length classes referring to four principle 5S-IGS types of disparate phylogenetic affinity.

While two (or more) paralogous (and/or homeologous) 5S loci may have facilitated ancient polymorphism (e.g. polyploidisation, hybridisation), intragenomic silencing of homeologues leading to pseudogeny (reviewed in Volkov et al., 2007; see Volkov et al., 2017 for a case of an ancient allopolyploid) may cause the observed detection differences. The HTS primers bind to the highly conserved 5S rDNA. If these are strongly degraded, the intergenic spacers of such arrays will not be in our sample. Compared to other studied Fagaceae (e.g. Denk & Grimm, 2010) and ITS data (Denk et al., 2005), reducing spacer length by reducing AT-dominated length-polymorphic regions may represent a general trend (root-tip distances in Fig. 5 reflect sequence derivation). Within beech, this trend is only obvious in the *crenata-sylvatica* lineage: the AT-richer A-types are slightly longer than the B-types replacing them. In *F. japonica*, longer, GC-richer O-types are rarer and less diverse than types I (Figs 6, 7; File S1, S2). Based on the high structural and sequence diversity, counterbalanced by the large number of identical sequences detected in each sample, a combined effect of both concerted and birth-and-death evolution models must therefore be assumed for the 5S rRNA genes in beech (Nei & Rooney, 2005; Galian et al., 2014). Thus, even if there are two (or more) loci in all species of *Fagus*, they may not be paralogous (in a strict sense) but rather act as homeologues, i.e. they differ in position but not in function and are affected by inter-array (inter-loci) recombination and limited concerted evolution.

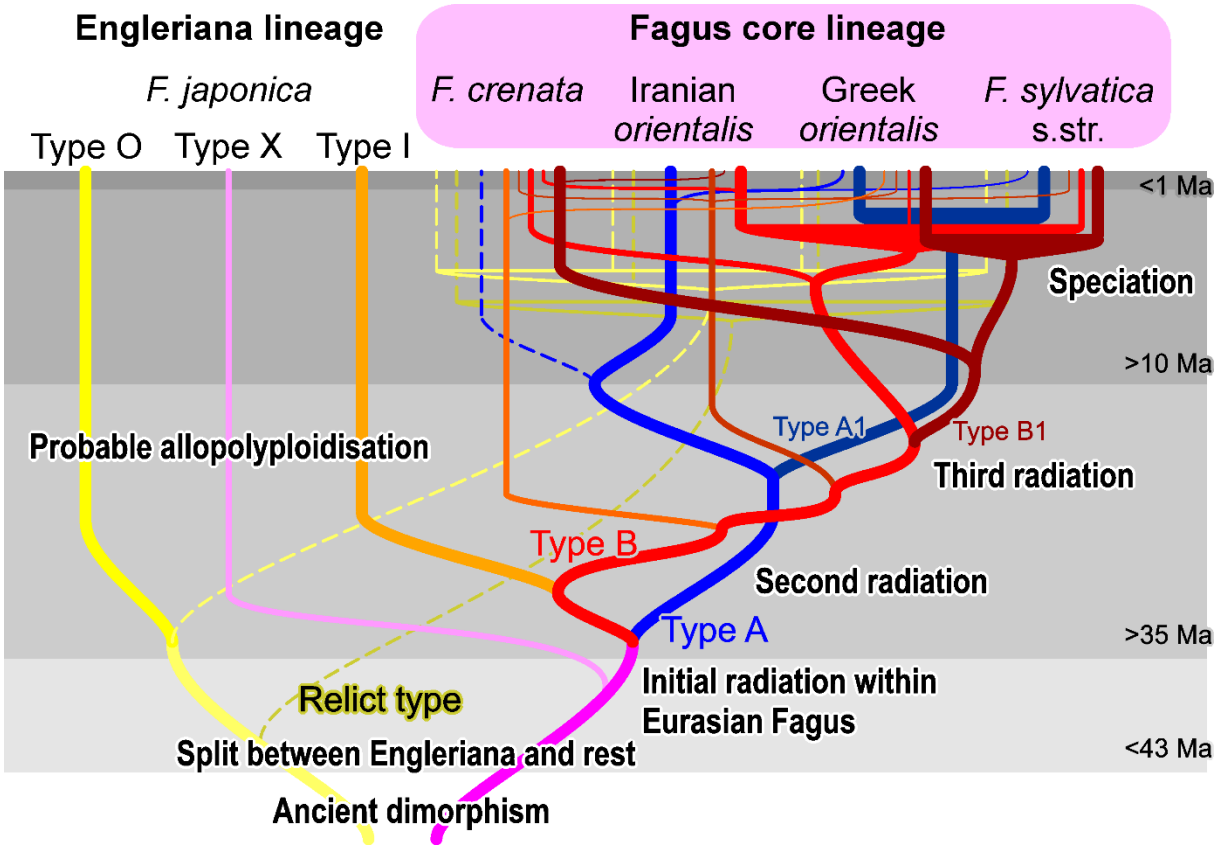


Figure 8 | Doodle summarising the totality of our results and their interpretation regarding all available (referenced in the text) information.

Formation of 5S-IGS gene pools in the F. crenata – F. sylvatica *s.str. lineage (Fig. 8)*

Based on the ML results of the 686-tip tree, the individual samples, and the 38 selected variants (Figs. 3, 5, 6), the potential disruption and assembly of ancient common gene pools can be discussed using the NNet network (Fig. 4). The closer two clusters (neighbourhoods, if defined by an edge bundle) are in the network, the more recent is their split. Mixed clusters typically comprise either shared ancestral variants or variants propagated during (past) gene flow. Pronounced 'trunks' imply genetic drift and isolation, while 'fans' reflect gradual radiation (see also Hipp et al., 2020 for oaks). The distances across the graph represent genetic distances, thus, the amount of genetic drift: the fixation rate of new mutations and their propagation within the genome and a population's gene pool. Further, the network does not assume dichotomy, which may be a poor model for the evolution and propagation of 5S-IGS variants within a species' gene pool. With currently available methods, it is impossible to date potentially paralogous-

homoeologous multi-copy 5S IGS data. Considering the used taxon sample, method and data, the divergence estimates by Renner et al. (2016) provide maximum ages for lineage splitting in beeches. 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). The fossil record indicates that gene flow between the western and eastern 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). 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). 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 type B' cluster is likely the direct result of a common origin 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). It is possible that *F. crenata* (the easternmost species) replaced its type A 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) (sub)species of beech 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 this area (and in 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, 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' lineages *(crenata-sylvatica* lineage).

In Europe, both A-type and B-type 5S-IGS arrays were secondarily sorted and homogenized to a certain degree: 'European A' and 'Western Eurasian B' clades include ± high proportion of shared ("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 the 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, which are likely relicts from a past diversification.

Inter-species relationships and status of Iranian F. orientalis

Our data confirm the close relationship of *F. crenata* with *F. sylvatica* s.l. and the deep split between the two Japanese species, belonging to different subgeneric lineages. Results also agree with (*i*) 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 Caucasus); and (*ii*) the relatively recent contact and mixing between the westernmost *F. orientalis* (here represented by a Greek sample) and *F. sylvatica* s.str. (Papageorgiou et al., 2008; Müller et al., 2019).

Considering all evidence (cf. Fig. 8), the Iranian *F. orientalis* must have been isolated for a much longer time and likely deserves recognition as distinct species. However, a formalisation requires comparative data of populations in the Caucasus (including NE. Turkey; cf. Gömöry and Paule, 2010), where the holotype of *F. orientalis* comes from. According to Denk (1999b), Gömöry & Paule (2010) and Gömöry et al. (2018), the Caucasian populations are morphologically and genetically distinct from both the western *F. orientalis* (including SE. Bulgarian and NW. Turkish populations) and the Iranian populations. The high amount and diversity of Iran-unique 5S-IGS A- and B-type variants, while being geographically restricted, points to a complex history of the Iranian beech. The Caucasian populations are geographically closer to the Iranian, and they may have been in contact in the more recent past (via Cis-Caucasia and Azerbaijan). Hence, some of the Iran-specific variants may be shared by the Caucasian populations. Further genetic imprints in the Iranian species could be a legacy from

extinct beech populations growing further east and the disjunct and taxonomically complex beeches from 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 above).

Potential of 5S-IGS HTS data to detect past and recent reticulation events

Beech has a complex biogeographic history in the Northern Hemisphere. The *crenata-sylvatica* lineage represents the most evolved, recently diverged branch (File S4; Denk & Grimm 2009; Renner et al., 2016). A reasonable model for beech evolution and diversification in western Eurasia would include phases of range contraction (isolation-speciation) and expansion (species mixing and homogenisation) to explain the diffuse morphological evolution in western Eurasian beeches. Such a scenario is supported by the observed 5S-IGS diversity (Fig. 8): speciation and isolation lead to the accumulation of new lineage-specific variants, which are then exchanged or propagated during episodes of favourable climate and massive radiation of beech forests. Without data from Chinese and/or Taiwanese species, it is impossible to assess whether the A- or B-types or the partly pseudogenic relict type lineage(s) represent the original stock of western Eurasian beeches and their Japanese sister species. Nonetheless, since the A-type is largely lost in *F. crenata*, and sequentially equally distant to the B-type than the *F. japonica* ingroup variants (I-type; Fig. 4), we postulate that both types are present at least in some of the Chinese species and represent an ancient polymorphism shared by all Eurasian 'subgenus Fagus' species.

Pronounced ancient nuclear polymorphism implies that at least some modern beeches are of hybrid or allopolyploid origin. For the species of 'subgenus Engleriana' a hybrid origin would make sense, with the *F. japonica* ingroup variants representing the common ancestry with the Eurasian clade of 'subgenus Fagus', while the outgroup variant represents another, potentially extinct, lineage of high-latitude beeches (Denk & Grimm, 2009; see also Fradkina et al., 2005; Grímsson et al., 2016). A hybrid/allopolyploid origin would also explain the extreme divergence observed in the ITS region of *Fagus* (Denk et al., 2005) and appears to be supported

by fossil evidence. Oldest fossils unambiguously belonging to *Fagus* (dispersed pollen grains from the early Paleocene of Greenland, with small size and 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 including morphological characteristics of both modern subgenera are also known. For example, the Eocene *Fagus langevinii* has branched cupule appendages as exclusively found in extant *Fagus 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 beeches, during which the modern subgenera were evolving, but not yet fully differentiated. By the early Oligocene, fossil-species can clearly be assigned to either 'subgenus Fagus' or 'Engleriana': *F. 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 *Fagus palaeojaponica*, *F. evenensis*, *Fagus* sp. 3).

Reticulation enriching the 5S-IGS pool 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. 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 type A, 'Original A' and 'European A'; shared 'Western Eurasian B'), we found a highly abundant, moderately evolved but less diverse B-type ('European B' in

Figs 3–6), exclusive to Greek *orientalis* and *F. sylvatica* s.str., likely reflecting the recent common origin of these populations. In the context of *F. gussonii*, the second fossil-species present in the Miocene of Europe (including Iceland) and of uncertain affinity (Denk & Grimm, 2009), data from North American extant species will be most valuable. Will they be clearly different (as found for the ITS region and 2nd *LEAFY* intron; Denk et al., 2005; Renner et al., 2016) or have they preserved affinities with one of the here established lineages?

Outlook

The evolution of extra-tropical tree genera involves dynamic speciation processes. *Fagus* species are wind-pollinated but animal-dispersed and have a narrow ecological niche. Therefore, beech populations are prone to becoming isolated during phases of area disruption. This can lead to speciation accompanied by lineage sorting. At the same time, species boundaries remain permeable: Fagaceae are known for both 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). Phases of area expansion likely led to secondary contacts and lineage mixing. Auto- and allopolyploidisation, hybridisation, introgression, and incomplete lineage sorting may leave complex genetic imprints. Single-copy or uniparentally inherited genetic markers can only capture some aspects of this complexity. Biparentally inherited, multi-copy, multi-locus markers, can instead provide direct evidence for past and recent reticulation. In this context, the 5S intergenic spacers are a unique source of information, being the only currently known region that (*i*) is highly divergent, (*ii*) can be easily amplified 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- and I/X-types found in *F. japonica* point towards a hybrid (allopolyploid) origin, and the distribution of A- and B-types in the *crenata-sylvatica* lineage reflect speciation processes predating the establishment of the modern species.

Beeches are ideal study objects because of the low species number and well-documented fossil record. Better understanding the diversity we observed and the putatively ancestral variants will require material from East Asian species not included in the present study: *F. engleriana, F. hayatae, F. longipetiolata,* and *F. lucida*, as well as from the yet poorly studied (eastern) North American *F. grandifolia* (but see Galván-Hernández et al., 2020). Similarly, nothing is known about 5S-IGS gene pools of East Asian and American oaks (and little about their ITS diversity); nevertheless, phylogenomic data of North American and French oaks have anticipated the existence of complex species relationships (Lepoittevin et al., 2015; Hipp et al., 2019). Non-trivial phylogenetic signals that may be the product of reticulate evolution also characterize all other Fagaceae (even complete plastomes cannot resolve the monophyly of oaks) and Fagales in general.

Further applications of 5S-IGS HTS data include taxonomy (co-)informed by genetics (e.g. Grimm & Denk, 2014; Liede-Schumann et al., 2020) and the assessment of global biodiversity in addition to the traditional counting of numbers of species (e.g. Forest et al., 2018). The species-status of *F. sylvatica* vs. *F. orientalis* has long been discussed but with focus on the wrong populations: while the western (here: Greek) *F. orientalis* may be addressed as a subspecies of *F. sylvatica*, the Iranian populations deserve the recognition as a species and represent a most valuable genetic resource.

Experimental procedure

Plant material and HTS methodology

Figure 1 shows our experimental design and workflow. Thirty beech individuals from three species (four taxa) were collected in the wild, covering 14 total populations. Four samples collect five individuals each of *Fagus sylvatica* s.l. from single populations in Germany, Italy (*F. sylvatica* s.str.), Greece, and Iran (*F. orientalis*); in addition, 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 and map provided in File S1, sections 1.2, 2.3).

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 prepared six artificial samples, consisting of pure species/taxa and/or populations, by pooling equal amounts of DNA of every individual to make up a total of 20 ng per sample. As in Piredda et al. (2021), the nuclear ribosomal 5S intergenic spacer (5S-IGS) was amplified with the plant-specific primer pair CGTGTTTGGGCGAGAGTAGT (forward) and CTGCGGAGTTCTGATGG (reverse). Paired-end Illumina sequencing (2×300 bp sequencing) was performed by LGC Genomics GmbH. 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 the differences in base calls 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. File 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 (File S1, section 2.3; File 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 visualisation 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, 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. In addition, we assessed intra- and inter-lineage diversity, between and within main 5S-IGS types (see *Results*), 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; File 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 (Cardoni et al., 2021, http://dx.doi.org/10.6084/m9.figshare.13793744.v1).

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. File S4 documents the selection and includes a fully annotated, tabulated version of the 38-tip alignment.

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), which financed the sequencing. 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.13793744.v1. Other relevant data are contained within the manuscript and its supporting materials (included in the same *figshare* file set: Cardoni et al., 2021; cf. *Supporting information*).

Supporting information (cf. Cardoni et al., 2021)

File 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 Supplementary file S2; appendix B includes violin and scatter plots sorted by sample and type.

File 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.

File 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*)

File 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.

References

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

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

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

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

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

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

Cardoni, S., Piredda, R., Denk, T., Grimm, G. W., Papageorgiou, A. C., Schulze, E.-D., Scoppola, A., Shanjani, P. S., Suyama, Y., Tomaru, N. Worth, J. R. P., & Simeone, M. C. Data for Cardoni et al. (2021): High-throughput sequencing of *Fagus* 5S-IGS rDNA. File set, *figshare*, doi:10.6084/m9.figshare.13793744.v1

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

Chokchaichamnankit, P., & Anamthawat-Jonsson, K. (2015). Cytogenetics of the rare and endangered *Trigonobalanus doichangensis* (Fagaceae) from Northern Thailand. *Journal of Tropical Forest Science*, 27, 60–68

Chokchaichamnankit, P., Anamthawat-Jonsson, K., & Chulalaksananukul, W. (2008). Chromosomal mapping of 18S–25S and 5S ribosomal genes on 15 species of Fagaceae from Northern Thailand. *Silvae Genetica*, 57, 5–13. doi:10.1515/sg-2008-0002

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

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

De Castro, O., Di Maio, A., Lozada García, J. A., Piacenti, D., Vázquez-Torres, M., & De Luca, P. (2013). Plastid DNA sequencing and nuclear SNP genotyping help resolve the puzzle of American *Platanus*. *Annals of Botany*, 112, 589–602. doi:10.1093/aob/mct134

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

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

Denk, T. (2003). Phylogeny of *Fagus* L. (Fagaceae) based on morphological data. *Plant Systematics and Evolution*, 240, 55–81. doi: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., Grimm, G., Stögerer, K., Langer, M., & 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. doi:10.1007/s006060200044

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

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

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

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

Ekblom, R., & Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, 107, 1–15. doi:10.1038/hdy.2010.152

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

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

Forest, F., & Bruneau, A. (2000). Phylogenetic analysis, organization, and molecular evolution of the non-transcribed spacer of 5S ribosomal RNA genes in *Corylus* (Betulaceae). *International Journal of Plant Science*, 161, 793–806. doi:10.1086/314294

Forest, F., Moat, J., Baloch, E., Brummitt, N. A., Bachman, S. P., Ickert-Bond, S., … Buerki, S. (2018). Gymnosperms on the EDGE. *Scientific Reports*, 8, 6053. doi:10.1038/s41598-018-24365-4

Forest, F., Savolainen, V., Chase, M. W., Lupia, R., Bruneau, A., & Crane, P. R. (2005). Teasing apart molecular- versus fossil-based error estimates when dating phylogenetic trees: A case study in the birch Family (Betulaceae). *Systematic Botany*, 30, 118–133. doi:10.1600/0363644053661850

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

Galián, J. A., Rosato, M., & Rossellò, J. A. (2014). Partial sequence homogenization in the 5S multigene families may generate sequence chimeras and spurious results in phylogenetic reconstructions. *Systematic Biology*, 63, 219–230. doi:10.1093/sysbio/syt101

Galván-Hernández, D. M., Octavio-Aguilar, P., Lazcano-Cruz, L., & Sánchez-González, A. (2020). Morphological and genetic differentiation in isolated populations of Mexican beech *Fagus grandifolia* subsp. *mexicana*. *Journal of Forest Research*, https://doi.org/10.1007/s11676-020-01247-y.

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

Göker, M., & Grimm, G. W. (2008). General functions to transform associate data to host data, and their use in phylogenetic inference from sequences with intra-individual variability. *BMC Evolutionary Biology*, 8, 86. doi:10.1186/1471-2148-8-86

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

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

Gömöry, D., Paule, L., & Vyšný, J. (2007). Patterns of allozyme variation in western Eurasian *Fagus*. *Botanical Journal of the Linnean Society*, 154, 165–174. doi:10.1111/j.1095-8339.2007.00666.x

Gouy, M., Guindon, S., & 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. doi:10.1093/molbev/msp259.

Grimm, G. W., & Denk, T. (2008). ITS Evolution in *Platanus* (Platanaceae): Homoeologues, Pseudogenes and Ancient Hybridization. *Annals of Botany*, 101, 403–419. doi:10.1093/aob/mcm305

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

Grimm, G. W., & Denk, T. (2012). Reliability and resolution of the coexistence approach - A revalidation using modern-day data. *Review of Palaeobotany and Palynology*, 172, 33–47. doi:10.1016/j.revpalbo.2012.01.006

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

Grimm, G. W., Denk, T., & 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. doi:10.1017/S1477200007002459

Grímsson, F., Grimm, G. W., Zetter, R., & 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. doi:10.1515/acpa-2016-0016

Hatziskakis, S., Papageorgiou, A. C., Gailing, O., & Finkeldey, R. (2009). High chloroplast haplotype diversity in Greek populations of beech (*Fagus sylvatica* L.) *Plant Biology*, 11, 425–433. doi: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. doi:10.1111/nph.16162

Hipp, A. L., Whittemore, A. T., Garner, M., Hahn, M., Fitzek, E., Guichoux, E., … Cannon, C. H. (2019). Genomic identity of white oak species in an eastern North American syngameon. *Annals of the Missouri Botanical Garden*, 104, 455–477. doi:10.3417/2019434

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. doi:10.1016/j.revpalbo.2019.02.004

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

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

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

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 15, 259–263. doi:10.1127/0941-2948/2006/0130

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & 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, 5112–5120. doi: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. doi:10.1101/gr.092759.109

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

Lamb, P. D., Hunter, E., Pinnegar, J. K., Creer, S., Davies, R. G., & Taylor, M. I. (2019). How quantitative is metabarcoding: A meta-analytical approach. *Molecular Ecology*, 28, 420–430. doi:10.1111/mec.14920

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

Lepoittevin, C., Bodénès, C., Chancerel, E., Villate, L., Lang, T., Lesur, I., … Kremer, A. (2015). Single-nucleotide polymorphism discovery and validation in high-density SNP array for genetic analysis in European white oaks. *Molecular Ecology Resources*, 15, 1446–1459. doi:10.1111/1755-0998.12407

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

Liede-Schumann, S., Grimm, G. W., Nürk, N. M., Potts, A. J., Meve, U., & Hartmann, H. E. K. (2020). Phylogenetic relationships in the southern African genus *Drosanthemum* (Ruschioideae, Aizoaceae). *PeerJ*, 8, e8999. doi:10.7717/peerj.8999

López de Heredia, U., Mora-Márquez, F., Goicoechea, P. G., Guillardín-Calvo, L., Simeone, M. C., & 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. doi:10.3389/fpls.2020.564414

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

Manchester, S. R., & 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. doi:10.1139/b04-112

McVay, J. D., Hipp, A. L., & 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. doi:10.1098/rspb.2017.0300

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

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

Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, 39, 121–152. doi:10.1146/annurev.genet.39.073003.112240

Oh, S.-H., Youm, J.-W., Kim, Y.-I., & 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. doi:10.1600/036364416X692271

Okaura, T., & Harada, K. (2002). Phylogeographical structure revealed by chloroplast DNA variation in Japanese beech (*Fagus crenata* Blume). *Heredity*, 88, 322–329. doi: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. doi:10.1007/s10342-007-0185-3

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

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

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11, 1633–1644. doi:10.5194/hess-11-1633-2007

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

Pilotti, M., Brunetti, A., Tizzani, L., & Marani, O. (2009). *Platanus x acerifolia* genotypes surviving to inoculation with *Ceratocystis platani* (the agent of canker stain): first screening and molecular characterization. *Euphytica*, 169, 1–17. doi:10.1007/s10681-009-9884-9

Piredda, R., Grimm, G. W., Schulze, E.-D., Denk, T., & 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. doi:10.1111/1755-0998.13264

Potts, A. J., Hedderson, T. A., & Grimm, G. W. (2014). Constructing phylogenies in the presence of Intra-Individual Site Polymorphisms (2ISPs) with a focus on the nuclear ribosomal cistron. *Systematic Biology*, 63, 1–16. doi:10.1093/sysbio/syt052

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. doi:10.1111/j.1469-8137.2011.03861.x

Renner, S. S., Grimm, G. W., Kapli, P., & 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. doi:10.1098/rstb.2015.0135

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

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. doi:10.1128/AEM.01541-09

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

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

Simeone, M. C., Cardoni, S., Piredda, R., Imperatori, F., Avishai, M., Grimm, G. W., & Denk, T. (2018). Comparative systematics and phylogeography of *Quercus* Section *Cerris* in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ*, 6, e5793. doi: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. doi:10.7717/peerj.1897

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

Volkov, R. A., Komarova, N. Y., & Hemleben, V. (2007). Ribosomal DNA in plant hybrids: inheritance, rearrangement, expression. *Systematics and Biodiversity*, 5, 261–276. doi:10.1017/S1477200007002447

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. doi:10.1186/s12870-017-0978-6

Wickham, H. (2016). ggplot2: elegant graphics for data analysis. Springer-Verlag, New York.

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. doi:10.1186/s12862-019-1523-z

Zhang, Z.-Y., Wu, R., Wang, Q., Zhang, Z.-R., Lopez-Pujol, J., Fan, D.-M., & 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. doi:10.1002/ece3.829