Systematics and Biodiversity **5** (3): 291–309
doi:10.1017/S1477200007002459 Printed in the United Kingdom

Issued 10 August 2007

Guido W. Grimm[1], Thomas Denk[2] & Vera Hemleben[1]*
[1]Department of General Genetics, Centre of Plant Molecular Biology (ZMBP), Eberhard Karls University, Auf der Morgenstelle 28, 72076 Tübingen, Germany
[2]Swedish Museum of Natural History, Department of Palaeobotany, Box 50007, 104 05 Stockholm, Sweden

submitted 12 December 2004
accepted 23 June 2005

Coding of intraspecific nucleotide polymorphisms: a tool to resolve reticulate evolutionary relationships in the ITS of beech trees (*Fagus* L., Fagaceae)

Abstract The internal transcribed spacers ITS1 and ITS2 of the nuclear ribosomal DNA (rDNA) have recently been found to display remarkable intraspecific polymorphism, a feature suggested as limiting their value for phylogenetic reconstructions. A comparative study of oligonucleotide motives and intraindividual nucleotide variability across all species of the tree genus *Fagus* (beech) shows, however, that this intraspecific ITS polymorphism follows a particular pattern, which can be used to detect reticulation and ancient polymorphism within the genus. Coding ITS polymorphisms as phylogenetically informative characters, moreover, resulted in better-resolved phylogenies than traditional 'base-per-base' maximum parsimony and maximum likelihood analyses.

Key words ITS, intraspecific variability, molecular evolution, phylogeny, reticulate evolution

Introduction

The internal transcribed spacers ITS1 and ITS2 of the nuclear ribosomal RNA genes (rDNA) are widely used molecular markers to infer low-level (intrafamiliar and intrageneric) phylogenetic relationships (Baldwin *et al.*, 1995; Jobst *et al.*, 1998; Álvarez & Wendel, 2003; Bailey *et al.*, 2003; Wissemann, 2003). With more data becoming available from numerous plant species, the genetic variability of the ITS has been shown to differ extremely between plant genera, although the overall length of the region comprising the ITS1, 5.8S rDNA, and ITS2 is similar in all angiosperms. For instance, in species of the genus *Acer* (Ackerly & Donoghue, 1998; Suh *et al.*, 2000; Tian *et al.*, 2002; Grimm, 2003) ITS1 and ITS2 are sufficiently distinct to allow straightforward phylogenetic analyses and taxonomic application. In contrast, species of *Fagus* (Stanford, 1998; Manos & Stanford, 2001; Denk *et al.*, 2002, 2005) are less differentiated and show a considerable amount of ambiguous sites at the intra- and interspecific level, including intragenomic variability (Denk *et al.*, 2002, 2005). To infer phylogenetic relationships from such a data set, commonly used methods such as distance methods, maximum parsimony, and maximum likelihood produce phylogenies that are problematic, since major divergence points lack crucial statistical support.

In the present paper we used the genus *Fagus* L. (beech, Fagaceae) as a model system to test the hypothesis that ITS polymorphisms yield additional information for, and are beneficial to, phylogenetic reconstructions. *Fagus* is a small genus comprising 10 tree species distributed throughout the northern hemisphere. Previous ITS studies did not entirely resolve intrageneric relationships (Stanford, 1998; Manos & Stanford, 2001; Denk *et al.*, 2002). Besides ITS1 and ITS2, other molecular markers such as cpDNA sequences were used, but infrageneric variability was too low to produce a reliable phylogeny for *Fagus* (Stanford, 1998). Denk *et al.* (2002) demonstrated that the intraspecific variability found within the ITS1 and ITS2 of the *F. sylvatica*-complex not only exceeds the overall generic variability, but causes serious problems for the usage of directly sequenced PCR products as data source and maximum parsimony (MP) as the analytical method. A maximum likelihood (ML) analysis carried out on the assembled data resulted in a phylogram that was in accordance with a previous morphological study (Shen, 1992), and recognized the two subgenera *Fagus* and *Engleriana* Shen. This was in conflict with Stanford's (1998) and Manos & Stanford's (2001) studies that used direct sequencing. Including all species of *Fagus* (Denk *et al.*, 2005; Table 1), a phylogeny was obtained that was highly congruent with a morphologically based phylogeny as well as with evidence from the fossil record (Fig. 1). The combined data (morphology, ITS, fossils) indicate that the high degree of ITS polymorphism in *Fagus* may be explained by the complex evolutionary behaviour of this molecular marker, the stenoecious ecological characteristic of *Fagus* in connection with its continuous geographical range through large parts of the Cenozoic, and by the absence of major radiations into diverse habitats.

*Corresponding author. Email: vera.hemleben@uni-tuebingen.de

Classification Shen (1992)	Occurrence
Fagus engleriana Seemen (incl. *F. multinervis* Nakai)	China, South Korea
Fagus japonica Maxim.	Japan
Fagus okamotoi Shen	Japan*
Fagus longipetiolata Seemen	China
Fagus brevipetiolata Hu	China*
Fagus bijiensis C.F. Wei & Y.T. Chang	western China†
Fagus tientaiensis T.N. Liou	eastern China†
Fagus lucida Rehder & Wilson	China
Fagus chienii Cheng	western China†
Fagus hayatae Palibin (incl. *F. pashanica* C.C. Yang)	China mainland, Taiwan
Fagus crenata Blume	Japan
Fagus sylvatica L. (incl. *F. orientalis* Lipsky, *F. moesiaca* (Maly) Czecz.)	Europe, southwestern Asia
Fagus grandifolia Ehrh. (incl. *F. mexicana* Martinez)	eastern North America, Mexico

Table 1 Taxonomy of *Fagus* (from Denk *et al.*, 2005).

*known from few localities, †known from a single one locality Species in normal script are not recognized as distinct species here and in previous studies by Denk (2003) and Denk *et al.* (2005)

Here we chose *Fagus* as a model group because of its high intraspecific ITS polymorphism. Since ML substitution models include only probabilities for point mutations, ambiguous data are treated as 'missing' or 'uncertain' and the phylogenetic information they contain is lost. To preserve this information, we code site variabilities and series of mutations as characters for a matrix that can be analysed either by MP or ML via Bayesian inference (BI).

Material and methods

Data source

We used the ITS data of Denk *et al.* (2005). ITS data were obtained after cloning and sequencing different PCR products representing all species of *Fagus* and including at least two individuals per species and several clones per individual (Denk *et al.*, 2005; accession numbers are listed in the Appendix, for further information refer to original literature). Gene bank sequences of other authors were not included, because they may lack crucial information about intraindividual variability due to the assembling procedure, i.e. direct sequencing of PCR products.

Alignment

The alignment of Denk *et al.* (2005) was used as the basis for the analyses. Basically, the alignment of *Fagus* ITS sequences

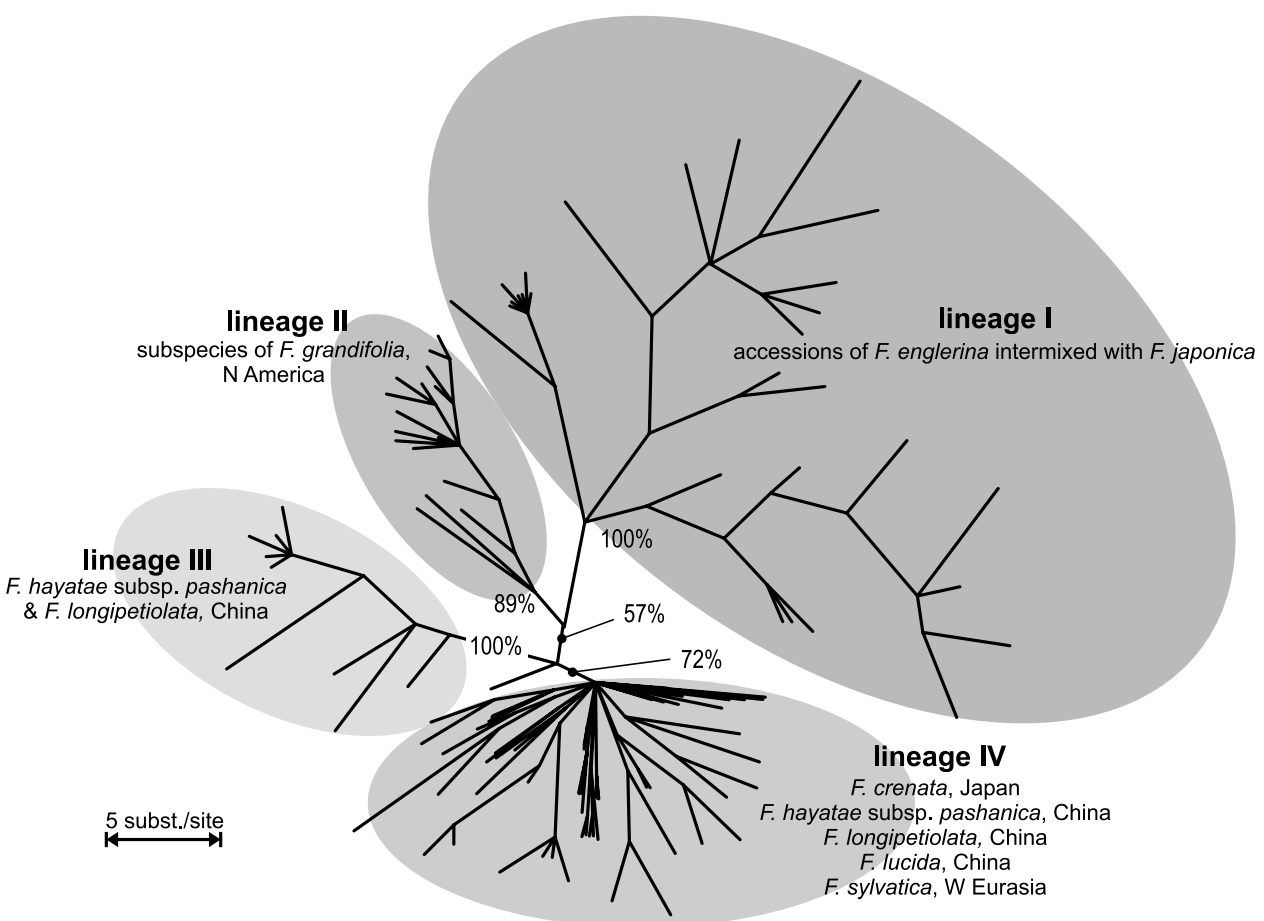

Figure 1 ML phylogram based on ITS data (after Denk *et al.*, 2005).

can be considered unbiased. The recognition of homologous nucleotide sites in the alignment was relatively easy because of the absence of common length polymorphism. Occasionally detected length polymorphisms (e.g. in the ITS1 of the subgenus *Engleriana* and in the ITS2 of *F. hayatae*) could be unambiguously located.

Coding of site variabilities

For the analyses the aligned sequence data were 'summarized' and transferred into matrices of characters (see below). Phylogenetic analyses were carried out with MrBayes 3.0 (Huelsenbeck & Ronquist, 2001, for ML) and PAUP® 4.0 beta 10 (Swofford, 2002, for MP). To minimize the number of taxa in the matrix, character states were summed up for species and geographical units within species. The coding follows a three-step protocol:

First step: Conversion of sequence data.

Second step: Assessing intraspecific variability.

Third step: Assembling and coding character states.

Conversion of sequence data (Table 2)

The conversion of sequence data is necessary primarily for oligonucleotide motives (see below) comprising 2 to 21 nucleotides. To maintain a coherent layout, single and linked site variabilities are equally transformed. In general, the nucleotide and/or oligonucleotide motif (i.e. a 'nucleotide state') found in the majority of all accessions is labelled as A_0. This must not be confused with a hypothetical ancestral state. Nucleotide states derived by point mutations from the consensus are labelled as B_0, C_0, etc., those derived through indels as X_1, X_2, where X = A, B, ... if derived from A_0, B_0, etc. For example, character 1 comprises three nucleotide states that can be ordered in a strictly sequential way and can be derived by a minimum of one fixed transition: CC (A_0) ↔ CT (B_0) ↔ TT (C_0). The derivation of differing nucleotide states (A_0, B_0, A_1, etc.) follows a parsimonious approach based on the fixed-character-state optimisation proposed by Wheeler (1999, 2001). Point mutations and indels are treated equally; i.e. an indel is considered to be the result of the fixation of a single mutational event. In case of equally parsimonious derivation pathways, substitution probabilities are taken into account, based on the ML via BI analysis by Denk *et al.* (2005). In the case of indels, deletions and insertions are favoured that can be attributed to simple duplications. For example, character 9 exhibits three nucleotide states: (i) C-5G, found in all Eurasian taxa of the subgenus *Fagus* (A_0); (ii) CA-4G (B_0), detected only in German individuals of *F. sylvatica*; (iii) C-3G (A_1) in *F. grandifolia*, subgenus *Engleriana*, and as intraindividual variability in *F. hayatae* subsp. *pashanica, F. longipetiolata* and *F. sylvatica* from Georgia. Both nucleotide states B_0 and A_1 can be derived from A_0 by a minimum of one fixed mutation, i.e. a transition (G ↔ A; A_0 ↔ B_0), and by deletion of two Gs (or the duplication of two Gs; A_0 ↔ A_1).

The nucleotide states of all clones of individuals of the same species are summarized in Table 2. Due to different numbers of individuals sampled per species and the range of the geographical area in which a species is found, accessions of some species are summarized in a different way: accessions of *F. engleriana* are divided into Chinese and South Korean specimens, and those of *F. grandifolia* into Mexican, south-eastern North American and eastern North American specimens corresponding to three subspecies recognized for *F. grandifolia* (Shen, 1992). *Fagus sylvatica* accessions are subdivided into geographical regions. For the present study, all occurring nucleotide states are taken into consideration, no matter whether they occur in only one, half of, or nearly all of the clones per individual.

Assessing intraspecific and intraindividual variability

Alignment sites containing differing nucleotide states (i.e. site variability) within a taxon (species, subspecies or individuals from the same geographical region) are treated as a character. Three basic types of site variabilities are recognized: (i) single site variabilities, (ii) oligonucleotide motives, and (iii) linked site variabilities:

Examples for single site variabilities are characters 2, 4, 5, 7, 8, 10, 12, 16, 19, 21, 23, 24, 27, 29, 30, 32, 34, 38, 39, 41, 42, 45–55, 57–60, and 62 (Table 2). In principle, single site variabilities may comprise up to 14 character states (depending on the number of possible combinations of nucleotides), however, only a maximum of four character states (site variabilities) is actually realized in species of *Fagus*. For example, at position 612 (character 46) each of the four nucleotides (A, C, G, and T) can be detected in clones of *Fagus*, while site variability is restricted for a particular taxon: the consensual G (state A_0) is found alone (no variability) or combined with A ($\{A_0B_0\}$ in subgenus *Engleriana*, C ($\{A_0C_0\}$ in *F. hayatae/longipetiolata*, or T ($\{A_0D_0\}$ in *F. sylvatica pro parte*.

Examples for oligonucleotide motives are characters 9, 14, 15, 18, 33, 36, 40 (Table 2). In case of oligonucleotide motives, site variability comprises length polymorphism and up to two point mutations. The detected site variabilities and the maximum parsimonious derivation of detected nucleotide states are in accordance (cf. Figs 2–4 in Denk *et al.*, 2005; see below).

Site variabilities at some positions within the ITS1 and ITS2 appear to be logically dependent, i.e. the nucleotide state at a certain site is consistently accompanied by the nucleotide state at another (nearby) site. Thus, a number of single site variabilities are not strictly independent in a parsimonious sense and are treated as one character. In our matrices, site variabilities are considered to be logically dependent or 'linked', if all clones with particular nucleotide states at one site show identical nucleotide types at the putatively linked (logically dependent) site. Linked single site variabilities can either be located next to each other such as in characters 1 (see above), 3, 11, 13, 17, 20, 22, 25, 26, 31, 35, 37, 43, 44, and 56 (Table 2), or separated by nucleotides and/or unlinked site variabilities (characters 6, 28, and 61; Table 2).

Assembling and coding character states (Table 3)

Variabilities encountered in distinct taxa and geographical areas (*F. engleriana*, *F. sylvatica*) are summed up and

				F. engleriana				*F. grandifolia*			*F. hayatae*			*F. sylvatica*				
	number of clones included in the analysis			China	Ullung Is., S.Korea	*F. japonica*	*F. crenata*	subsp. *grandifolia*	subsp. *carolinia-ana*	subsp. *mexicana*	subsp. *pashanica*	*F. longi-petiolata*	*F. lucida*	Georgia	Turkey	Hungary/ Slovenia	Germany	Italy/ Spain
				16	5	8	9	7	5	4	13	16	8	7	10	8	9	16
character	alignment sites	consensus (A_o-state)	other states															
1	78,79	CC	**T**C=B_o T**T**=C_o	{B_oC_o}	{B_oC_o}	{B_oC_o}	A_o	{A_oB_o}	{A_oB_o}	A_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	{A_oB_o}	{A_oB_o}
2	87	C	**T**=B_o	A_o	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o	{A_oB_o}	A_o	A_o	A_o	{A_oB_o}	{A_oB_o}	A_o
3	98–100	CGC	CA**A**C=B_o C**T**C=C_o CG**A**=D_o T**TA**=E_o T**T**GC=F_o	{$A_oD_oE_o$}	{$A_oD_oE_o$}	{$A_oD_oE_o$}	A_o	{B_oC_o}	C_o	{B_oC_o}	A_o	A_o	A_o	{A_oC_o}	A_o	{A_oF_o}	A_o	A_o
4	108	C	**G**=B_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
5	121	G	**A**=B_o **C**=C_o	A_o	A_o	A_o	{A_oB_o}	A_o	A_o	A_o	X_o	X_o	{A_oC_o}	{A_oB_o}	A_o	A_o	{A_oB_o}	{A_oB_o}
6	126,128	T...C	**C**...C=B_o T...**T**=C_o	A_o	A_o	A_o	{A_oC_o}	A_o	A_o	A_o	A_o	A_o	{A_oC_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	{A_oB_o}
7	135	T	**C**=B_o	A_o	A_o	A_o	A_o	B_o	{A_oB_o}	B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
8	139	G	**A**=B_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
9	152–157	CGGGGG	CA**G**GGG=B_o CGGG**xx**=A_1	A_1	A_1	A_1	A_o	A_1	A_1	A_1	{A_oA_1}	{A_oA_1}	A_o	{A_oA_1}	A_o	A_o	{A_oB_o}	X_o
10	159	A	**G**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o
11	162–165	CCGT	**T**CGT=B_o **A**CGT=C_o C**T**GT=D_o CCG**C**=E_o	{$A_oB_oE_o$}	{A_oE_o}	{A_oE_o}	A_o	A_o	{A_oC_o}	A_o	{$A_oB_oD_o$}	{A_oD_o}	A_o	{$A_oB_oC_o$}	A_o	A_o	A_o	A_o
12	167	C	**T**=B_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o	A_o	{A_oB_o}	A_o	{A_oB_o}	A_o
13	171,172	CC	**T**C=B_o **G**C=C_o C**T**=D_o C**A**=E_o	A_o	A_o	{$A_oD_oE_o$}	{A_oD_o}	A_o	A_o	{A_oC_o}	A_o	A_o	A_o	A_o	{A_oB_o}	A_o	{A_oD_o}	{$A_oB_oD_o$}
14	180–185	CACAAA	**xx**CAAA=A_1 C**G**CAAA=B_o CACA**G**A=C_o	{A_oA_1}	{A_oA_1}	{A_oA_1}	{$A_oB_oA_1$}	{A_oC_o}	{A_oC_o}	{A_oC_o}	A_o	A_o	A_o	{A_oA_1}	A_o	A_o	A_o	A_o
15	187–207	short type	long type=B_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
16	212	G	**A**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o
17	216,217	GT	**A**T=B_o G**C**=C_o	{$A_oB_oC_o$}	{A_oC_o}	A_o	A_o	A_o	A_o	A_o	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	A_o
18	220–226	CAACC	CAA**AA**CC=A_1 **T**AACC=B_o **A**AACC=C_o AAA**T**C=D_o CAA**A**C=E_o **G**AA**G**C=F_o	{A_oD_o}	{A_oD_o}	{A_oD_o}	A_o	A_o	A_o	A_o	A_o	{A_oA_1}	{A_oC_o}	A_o	{A_oE_o}	{$A_oB_oF_o$}	{A_oF_o}	A_o
19	228	C	**T**=B_o **G**=C_o	{A_oB_o}	{A_oB_o}	{A_oB_o}	A_o	A_o	A_o	A_o	A_o	{A_oC_o}	A_o	A_o	A_o	A_o	A_o	A_o
20	233,234	GT	**A**T=B_o G**C**=C_o G**A**=D_o	A_o	{A_oD_o}	A_o	{A_oB_o}	{A_oC_o}	A_o	A_o	A_o	A_o	{A_oC_o}	A_o	A_o	A_o	{A_oB_o}	A_o

Table 2 Detected nucleotide states summarized for species and geographic areas.

				F. engleriana				*F. grandifolia*			*F. hayatae*			*F. sylvatica*				
				China	Ullung Is., S.Korea	*F. japonica*	*F. crenata*	subsp. *grandifolia*	subsp. *carolinia-ana*	subsp. *mexicana*	subsp. *pashanica*	*F. longi-petiolata*	*F. lucida*	Georgia	Turkey	Hungary/ Slovenia	Germany	Italy/ Spain
	number of clones included in the analysis			16	5	8	9	7	5	4	13	16	8	7	10	8	9	16
character	alignment sites	consensus (A_o-state)	other states															
21	240	G	**T**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o
22	249,250	CG	**T**G=B_o C**A**=C_o	$\{A_oC_o\}$	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
23	269	T	**C**=B_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
24	275	C	**A**=B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
25	284–286	TCG	**C**CG=B_o T**A**G=C_o TC**A**=D_o	X_o/B_o[†]	$\{A_oD_o\}$	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o
26	291,292	CC	**T**C=B_o C**T**=C_o	A_o	A_o	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o
27	294	C	**A**=B_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o
28	306,318	T...C	T...**T**=B_o **A**...T=C_o	X_o	X_o	X_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
29	310	C	**T**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$
30	316	G	**A**=B_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o
31	505.506	CC	**T**C=B_o T**T**=D_o C**T**=C_o	X_o[‡]	$\{B_oD_o\}$	X_o/C_o	$\{A_oB_o\}$	B_o	B_o	B_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	$\{A_oC_o\}$	X_o/D_o
32	512	C	**T**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o
33	521–525	CCCC	CC**T**CC=A_1 CCC**T**=C_o	C_o	C_o	C_o	A_o	$\{A_oA_1\}$	$\{A_oA_1\}$	A_o	$\{A_oC_o\}$	$\{A_oC_o\}$	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o
34	527	G	**A**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o
35	531–533	CGC	**T**GC=B_o C**A**C=C_o CG**T**=D_o	$\{A_oC_o\}$	$\{A_oC_o\}$	$\{A_oC_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	$\{A_oB_o\}$	X_o/C_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o
36	534–538	gap	CTCCC insert=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o
37	544–548	GCGCG	**C**CGCG=B_o G**T**GCG=C_o GC**C**CG=D_o GC**T**CG=E_o GCG**T**G=F_o GCGC**C**=G_o	$\{A_oG_o\}$	$\{A_oG_o\}$	$\{A_oG_o\}$	A_o	A_o	A_o	A_o	$\{A_oC_oD_oF_oG_o\}$	$\{A_oB_oF_o\}$	$\{A_o\ B_o\ C_oD_o\}$	A_o	$\{A_oC_oD_oE_o\}$	A_o	$\{A_oC_o\}$	$\{A_oC_oE_o\}$
38	552	T	**G**=B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
39	555	C	**T**=B_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o
40	562–565	TGG	**C**GG=B_o CGG**G**=B_1 TG**A**=C_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	X_o/B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	$\{A_oC_o\}$	A_o
41	567	G	**A**=B_o **T**=C_o **C**=D_o	X_o	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o

Table 2 Continued.

				F. engleriana				*F. grandifolia*			*F. hayatae*							
	number of clones included in the analysis			China	Ullung Is., S.Korea	*F. japonica*	*F. crenata*	subsp. *grandifolia*	subsp. *caroliniana*	subsp. *mexicana*	subsp. *pashanica*	*F. longipetiolata*	*F. lucida*	Georgia	Turkey	Hungary/ Slovenia	Germany	Italy/ Spain
				16	5	8	9	7	5	4	13	16	8	7	10	8	9	16
character	alignment sites	consensus (A_o-state)	other states															
42	585	G	**T**=B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o
43	588–591	CTGT	**T**GTG=B_o C**G**GT=C_o CT**A**T=G_o CT**A**C=E_o CTG**C**=D_o CTG**G**=F_o	$\{A_oB_oD_oE_oF_o\}$	$\{A_oD_oE_o\}$	$\{A_oB_oD_oE_o\}$	A_o	A_o	A_o	A_o	A_o	$\{A_oD_o\}$	A_o	A_o	$\{A_oC_oG_o\}$	A_o	A_o	A_o
44	595,596	CG	**T**G=B_o C**A**=C_o	A_o	A_o	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	X_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o
45	602	A	**G**=B_o	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o
46	612	G	**A**=B_o **C**=C_o **T**=D_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	$\{A_oC_o\}$	$\{A_oC_o\}$	A_o	$\{A_oD_o\}$	A_o	$\{A_oD_o\}$	A_o	A_o
47	619	T	**C**=B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	$\{A_oB_o\}$
48	622	G	**A**=B_o **T**=C_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o	A_o
49	626	C	**T**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o
50	653	T	**G**=B_o	B_o	B_o	B_o	A_o	B_o	B_o	B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o
51	671	G	**A**=B_o	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o
52	674	C	**T**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o
53	676	T	**C**=B_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o
54	679	C	**T**=B_o	$\{A_oB_o\}$	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
55	686	C	**T**=B_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o
56	689–691	CAA	CA**C**=B_o **T**AC=C_o C**G**C=D_o	$\{B_oC_o\}$	$\{B_oC_o\}$	$\{B_oC_o\}$	$\{A_oB_o\}$	B_o	B_o	B_o	$\{A_oC_o\}$	X_o/D_o	$\{A_oB_o\}$	$\{A_oB_o\}$	X_o/C_o	X_o/C_o	$\{A_oD_o\}$	X_o/C_o
57	702	G	**A**=B_o **T**=C_o	A_o	A_o	A_o	X_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oC_o\}$	A_o	A_o	A_o	A_o	A_o
58	704	C	**T**=B_o	$\{A_oB_o\}$	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
59	709	C	**T**=B_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o
60	713	C	**T**=B_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	A_o	A_o	A_o	A_o	A_o	A_o
61	716–724	G...C	G...**A**=B_o **A**...C=C_o	$\{B_oC_o\}$	$\{B_oC_o\}$	$\{B_oC_o\}$	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o	A_o
62	736	A	**G**=B_o	A_o	A_o	A_o	A_o	$\{A_oB_o\}$	$\{A_oB_o\}$	B_o	A_o	A_o	A_o	$\{A_oB_o\}$	A_o	A_o	A_o	A_o

Table 2 Continued.

† X_o/Y_o = all nucleotide states except for Y_o realised.
‡ X_o = all nucleotide states realised.

character	type	coding
1	ordered	A_o=a {A_oB_o}=b {B_oC_o}=d
2	binary	A_o=a {A_oB_o}=b
3	complex	A_o=a {A_oC_o}=b C_o=c {B_oC_o}=d {$A_oD_oE_o$}=e {A_oF_o}=f
4	binary	A_o=a {A_oB_o}=b
5	complex	A_o=a {A_oB_o}=b {A_oC_o}=c X_o*=d
6	ordered	{A_oC_o}=a A_o=b {A_oB_o}=c
7	ordered	A_o=a {A_oB_o}=b B_o=c
8	binary	A_o=a {A_oB_o}=b
9	complex	{A_oB_o}=a A_o=b {A_oA_1}=c A_1=d X_i=e
10	binary	A_o=a {A_oB_o}=b
11	complex	A_o=a {A_oC_o}=b {A_oD_o}=c {A_oE_o}=d {$A_oB_oC_o$}=e {$A_oB_oD_o$}=f {$A_oB_oE_o$}=g
12	binary	A_o=a (A_oB_o)=b
13	complex	A_o=a {A_oB_o}=b {A_oC_o}=c {A_oD_o}=d {$A_oB_oD_o$}=e {$A_oD_oE_o$}=f
14	complex	A_o=a {A_oA_1}=b {$A_oB_oA_1$}=c {A_oC_o}=d
15	binary	A_o=a {A_oB_o}=b
16	binary	A_o=a {A_oB_o}=b
17	complex	A_o=a {A_oB_o}=b {A_oC_o}=c X_o=d
18	complex	A_o=a {A_oA_1}=b {A_oC_o}=c {A_oD_o}=d {A_oE_o}=e {A_oF_o}=f {$A_oB_oF_o$}=g
19	ordered	{A_oB_o}=a A_o=b {A_oC_o}=c
20	complex	A_o=a {A_oB_o}=b {A_oC_o}=c {A_oD_o}=d
21	binary	A_o=a {A_oB_o}=b
22	ordered	{A_oB_o}=a A_o=b {A_oC_o}=c
23	ordered	A_o=a {A_oB_o}=b B_o=c
24	binary	A_o=a {A_oB_o}=b
25	complex	A_o=a {A_oB_o}=b {A_oC_o}=c {A_oD_o}=d {$A_oC_oD_o$}=e
26	ordered	{A_oB_o}=a A_o=b {A_oC_o}=c
27	binary	A_o=a {A_oB_o}=b
28	ordered	A_o=a X_o=c
29	binary	A_o=a {A_oB_o}=b
30	binary	A_o=a {A_oB_o}=b
31	complex	A_o=a {A_oB_o}=b B_o=c {B_oD_o}=d {$A_oB_oD_o$}=e X_o=f {A_oC_o}=g {$A_oB_oC_o$}=h
32	binary	A_o=a {A_oB_o}=b
33	ordered	C_o=a {A_oC_o}=b A_o=c {A_oA_1}=d
34	binary	A_o=a {A_oB_o}=b
35	complex	A_o=a {A_oB_o}=b {A_oC_o}=c {$A_oB_oD_o$}=d
36	binary	A_o=a {A_oB_o}=b
37	complex	A_o=a {$A_oB_oF_o$}=b {A_oC_o}=c {$A_oB_oC_oD_o$}=d {$A_oC_oD_oE_o$}=e {$A_oC_oD_oF_oG_o$}=f {A_oG_o}=g {$A_oC_oE_o$}=h
38	binary	A_o=a {A_oB_o}=b
39	binary	A_o=a {A_oB_o}=b
40	complex	A_o=a {A_oB_o}=b {A_oC_o}=c {$A_oB_1C_o$}=d
41	complex	A_o=a {A_oB_o}=b {A_oC_o}=c X_o=d
42	binary	A_o=a {A_oB_o}=b
43	complex	A_o=a {$A_oB_oD_oE_o(F_o)$}=b {$A_oD_oE_o$}=c {$A_oC_oG_o$}=d {A_oD_o}=e
44	complex	A_o=a {A_oB_o}=b {A_oC_o}=c X_o=d
45	binary	A_o=a {A_oB_o}=b
46	complex	A_o=a {A_oB_o}=b {A_oC_o}=c {A_oD_o}=d
47	binary	A_o=a {A_oB_o}=b
48	ordered	{A_oB_o}=a A_o=b {A_oC_o}=c
49	binary	A_o=a {A_oB_o}=b
50	ordered	A_o=a {A_oB_o}=b B_o=c
51	binary	A_o=a {A_oB_o}=b
52	binary	A_o=a {A_oB_o}=b
53	binary	A_o=a {A_oB_o}=b
54	binary	A_o=a {A_oB_o}=b
55	binary	A_o=a {A_oB_o}=b
56	complex	A_o=a {A_oB_o}=b B_o=c {A_oC_o}=d {$A_oB_oC_o$}=e {B_oC_o}=f {A_oD_o}=g {$A_oB_oD_o$}=h
57	complex	A_o=a {A_oB_o}=b {A_oC_o}=c X_o=d
58	binary	A_o=a {A_oB_o}=b
59	binary	A_o=a {A_oB_o}=b
60	binary	A_o=a {A_oB_o}=b
61	ordered	A_o=a {B_oC_o}=d
62	ordered	A_o=a {A_oB_o}=b B_o=c

Table 3 Character coding.

* X_o, all nucleotide states realized.

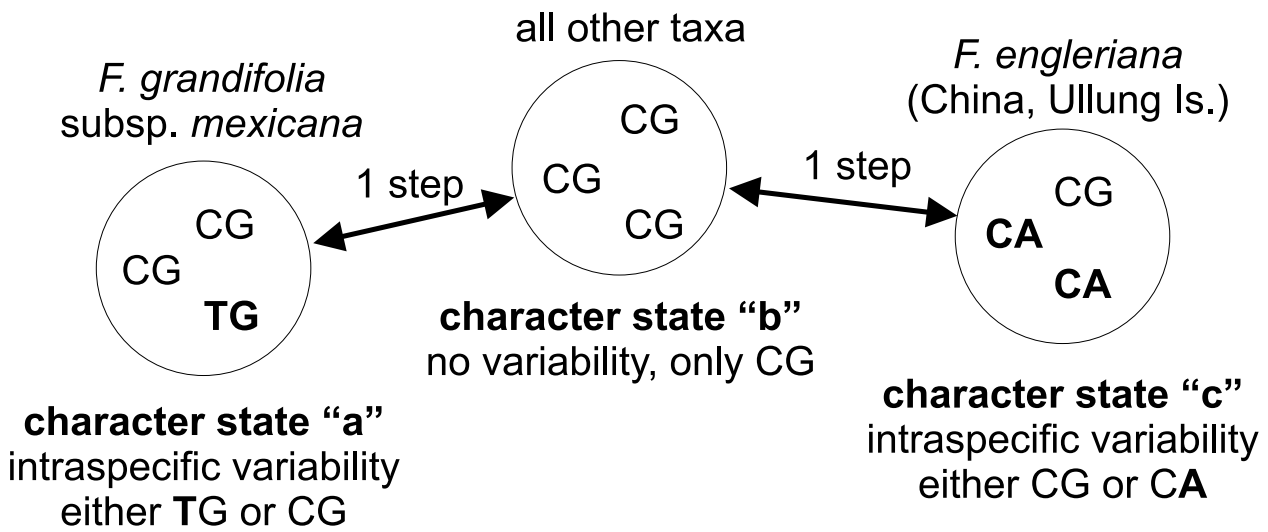

Figure 2 Example for a three-state ordered character (character 22; Tables 2, 3).

transformed into matrices. The coding is based on the presence and/or absence of a certain site variability/-ies. If some taxa show a nucleotide state A_0 (e.g. C), others a state B_0 (e.g. T), and the rest either A_0 or B_0, the matrices contain three character states: $a = A_0$, no variability (i.e. all clones of a taxon exhibit a C); $b = \{A_0B_0\}$, genetic variability is preserved in different populations/accessions (i.e. clones with C or T); and $c = B_0$, no variability (all clones with T). A character contains up to eight possible states and comprises either a single site or a number of sites, i.e. an oligonucleotide motif of logically dependent sites (see above). Characters in the data matrix meet the requirements for 'good' parsimonious informative characters, i.e. they are independent from each other and distinguishable (cf. Forey *et al.*, 1992, and references cited there).

Three character types are distinguished: (i) binary characters, (ii) ordered characters and (iii) complex characters.

Binary characters are defined by the occurrence or lack of one type of site variability: $a = A_0$, no variability; $b = \{A_0B_0\}$, site variability. This is the most common type of site variability within the ITS of *Fagus*: taxa are distinguished by displaying only the consensual nucleotide state (e.g. C), or clones with the consensual state and a state derived by a single transition (i.e. C or T). Binary characters are treated as unordered in all analyses.

Ordered characters comprise (i) characters with two different nucleotide states and the resulting site variability (e.g. character 7: $a = A_0$ in most *Fagus*, $b = \{A_0B_0\}$ in *F. grandifolia* subsp. *caroliniana*, and $c = B_0$ in the remaining subspp. of *F. grandifolia*; Table 2) and (ii) characters with two or more variabilities. An example for a simple ordered character with two variabilities and the underlying nucleotide states is given in Fig. 2 based on character 22 (positions 249, 250: $a = \{A_0B_0\} =$ TG or CG, $b = A_0 =$ CG, and $c = \{A_0C_0\} =$ CG or CA). Ordered characters can also be defined based on more complex variability patterns if a clear derivation of nucleotide states and detected site variability is possible as in character 1 (see above). Here, the actually detected variabilities, i.e. no variability (CC, a) ↔ CC/CT variability (b) ↔ CT/TT variability (c) are in accordance with the most parsimonious mutation sequence at the nucleotide level: CC ↔ CT ↔ TT.

Ordered characters are defined as 'ordered' in the first analysis (see Results), i.e. it takes two steps from 'a' to 'c' for the given examples under MP. Under the ML method, the direct transformation of 'a' to 'c' is prohibited by setting the character type to ordered.

Characters comprising a number of possible variabilities are coded as complex characters. The derivation of character states is modelled by a step-matrix. Within the step-matrix each gain or loss of individual variabilities equals one step. A very simple step-matrix (three competing site variabilities + lack of variability) is defined in the 'div_var' step-matrix (Fig. 3). In the example given above for the single site character 46 (either A, C, G, or T at pos. 612) the consensual nucleotide (state A_0) appears alone (character state 'a') or together with one of the remaining nucleotides ($\{A_0B_0\}$,$\{A_0C_0\}$,$\{A_0D_0\}$; character states 'b', 'c', and 'd'). In analogy to binary characters, a single step is required for 'a' ↔ 'b' ('c' and 'd', respectively; Fig. 3). Step matrices are further required in the following cases: (i) A character state can be derived from two different states. (ii) A character state is defined by the combination of site variabilities (character states). For example, at position 121 (character 5) clones of *F. hayatae* and *F. longipetiolata* exhibit either the consensual G, A or C (character state 'd'; Tables 3, 4). G or A ('b') is found as site variability in *F. crenata* and *F. sylvatica*, whereas *F. lucida* exhibits G or C ('c'). Other taxa do not show site variability ('a'). One step is required from 'a' (no variability) to 'b' and 'c' (site variability). Hence, 2 steps are required from 'b' to 'c' (cf. Fig. 2), and two steps from 'a' to 'd' (gained 2 variabilities: 'b' and 'c'). One step is again required from 'b' or 'c' to 'd'. More complex step-matrices are applied in case of linked site variabilities (Fig. 4). Step-matrices for all complex characters used in the analyses are provided in the appendix. Since ML is a process-, not character-based analysing method, step-matrices cannot be used. Instead, characters coded as step-matrices are either treated as unordered characters or divided into a group of binary/ordered characters. The resulting matrices are given in Table 4.

For each method (MP and ML) two runs were performed: Firstly, MP and ML analyses were carried out taking

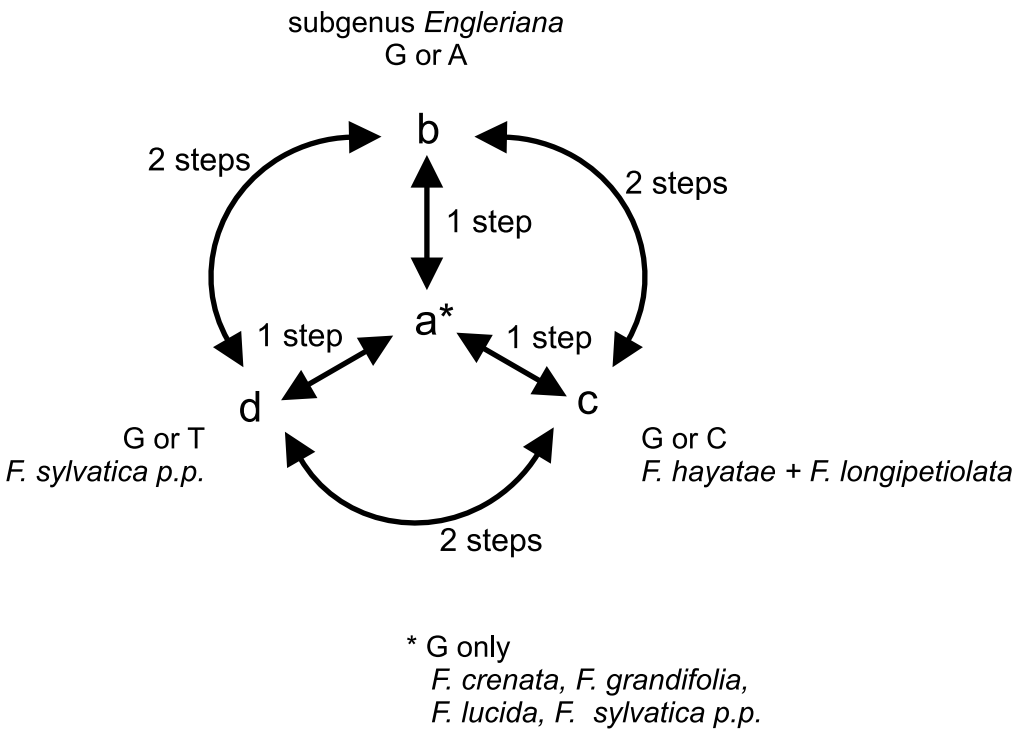

Figure 3 The DIV_VAR step-matrix, an example for a complex character with four character states (exemplarily illustrated for character 45; Tables 2, 3). Character state 'a': no intrataxonomic variability detected; character states 'b', 'c', and 'd': different intrataxonomic variabilities found.

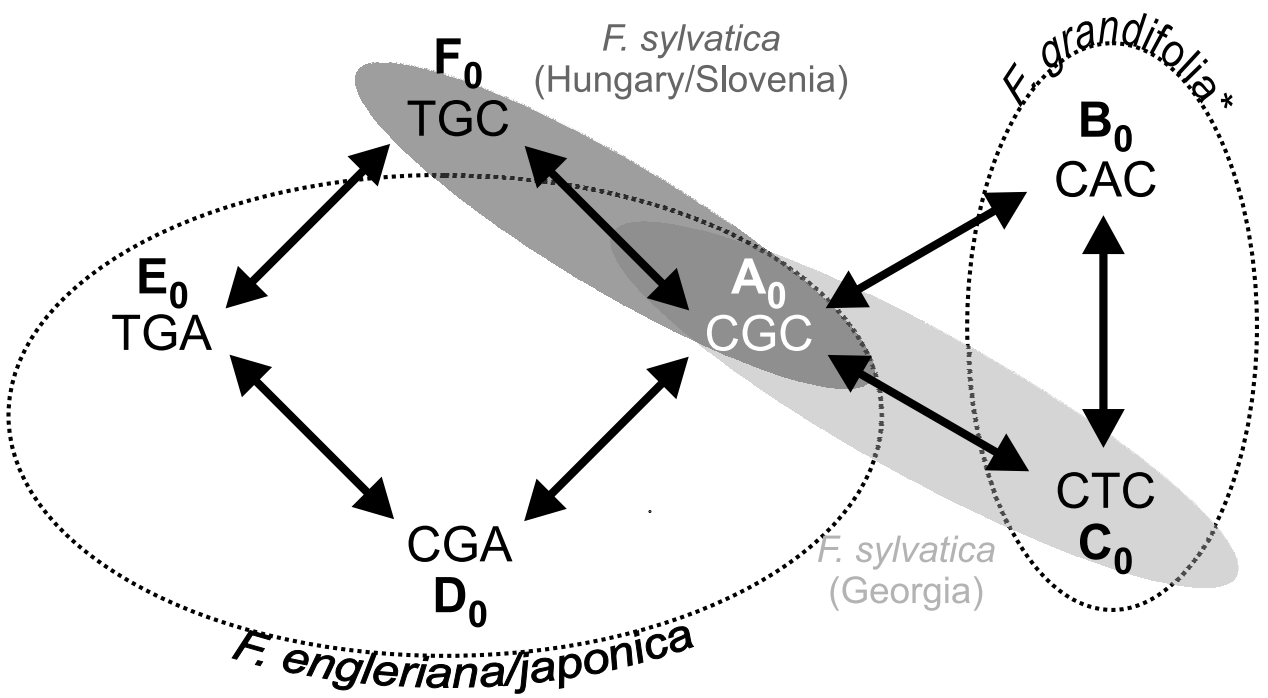

Figure 4 Linked site variability and step-matrix coding (character 3, Table 2). Although all three sites included show point mutations, the detected intraindividual/-specific variabilities comprise only such nucleotide states ('A_o', etc.) that can be directly derived from each other by single (not concerted) point mutations. Moreover, particular mutations appear to be linked (see text).
* In clones of *F. grandifolia* subsp. *caroliniana* only state C_o is found.

		ITS1																												
	character	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
F. engleriana	China	d	a	e	b	a	b	a	b	d	a	g	b	a	b	b	a	d	d	a	a	a	c	a	b	e	b	a	c	a	a
	Ullung	d	b	e	b	a	b	a	b	d	a	d	b	a	b	b	a	c	d	a	d	a	c	a	b	d	b	a	c	a	a
F. japonica		d	a	e	b	a	b	a	b	d	a	d	b	f	b	b	a	a	d	a	a	a	b	a	b	c	c	a	c	a	a
F. crenata		a	a	a	a	b	a	a	a	b	a	a	b	d	c	a	a	a	a	b	b	a	b	a	a	a	b	a	a	a	b
F. grandifolia	subsp. *grandifolia*	a	a	c	a	a	b	c	a	d	a	a	a	a	d	a	a	a	a	b	c	a	b	c	a	a	b	a	a	a	a
	subsp. *caroliniana*	b	a	c	a	a	b	b	a	d	a	b	a	a	d	a	a	a	a	b	a	a	b	b	a	a	b	b	a	a	a
	subsp. *mexicana*	a	a	d	a	a	b	c	a	d	a	a	a	c	d	a	a	a	a	b	a	a	a	c	a	a	b	a	a	a	a
F. hayatae	subsp. *pashanica*	b	a	a	a	d	b	a	a	c	b	f	a	a	a	a	b	b	a	b	a	a	b	a	a	a	b	a	a	a	a
F. longipetiolata		b	b	a	a	d	b	a	a	c	b	c	a	a	a	a	b	a	b	c	a	a	b	a	a	b	b	a	a	a	b
F. lucida		b	a	a	a	c	a	a	a	b	b	a	a	a	a	a	a	a	c	b	c	a	b	a	a	b	b	a	a	a	a
F. sylvatica	Georgia	b	a	b	a	b	c	a	a	c	a	e	a	a	b	a	a	a	a	b	a	b	b	a	a	a	a	a	a	a	b
	Turkey	a	a	a	a	a	c	a	a	b	a	a	b	b	a	a	a	a	e	b	a	b	b	a	a	a	a	b	a	a	a
	Hungary, Slovenia	a	b	f	a	a	b	a	a	b	a	a	a	a	a	a	a	a	g	b	a	a	b	a	a	a	b	a	a	a	a
	Germany	b	b	a	a	b	b	a	a	a	a	a	b	d	a	a	a	a	f	b	b	a	b	a	a	a	b	a	a	a	a
	Italy, Spain	b	a	a	a	b	b	a	a	e	a	a	a	e	a	a	a	a	a	b	a	a	b	a	a	a	b	a	a	b	a

		ITS2																														
	character	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**	**61**	**62**
F. engleriana	China	f	a	a	a	c	a	g	b	b	a	d	b	b	a	b	b	b	a	a	c	b	a	a	b	a	f	a	b	a	b	d	a
	Ullung	d	a	a	a	c	a	g	b	a	a	c	b	c	a	a	b	b	a	a	c	a	a	a	a	a	f	a	b	b	a	d	a
F. japonica		e	a	a	a	c	a	g	b	a	a	a	b	b	c	b	b	b	b	a	c	b	a	a	b	a	f	a	b	a	a	d	a
F. crenata		b	a	c	a	b	a	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	b	a	a	b	d	a	a	a	a	a
F. grandifolia	subsp. *grandifolia*	c	a	d	a	a	a	a	a	a	b	a	a	a	a	a	a	a	b	a	c	a	a	a	a	a	c	a	a	a	a	a	c
	subsp. *caroliniana*	c	a	d	a	a	a	a	a	a	b	a	a	a	a	a	a	a	b	a	c	a	a	a	a	a	c	a	a	a	a	a	b
	subsp. *mexicana*	c	a	c	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a	c	a	a	a	a	a	c	a	a	a	a	a	c
F. hayatae	subsp. *pashanica*	a	a	b	a	b	b	f	a	a	d	a	b	a	a	b	c	b	b	a	b	a	a	b	a	a	d	a	a	a	b	a	a
F. longipetiolata		b	b	b	a	d	a	b	a	b	b	b	b	e	a	a	c	b	c	a	b	b	a	b	a	a	e	b	a	a	b	a	a
F. lucida		b	b	c	a	b	a	d	a	a	b	a	a	a	d	a	a	a	b	b	a	a	a	b	a	a	b	c	a	a	a	a	a
F. sylvatica	Georgia	b	a	c	a	a	a	a	a	a	a	b	a	a	a	a	d	b	b	a	b	a	b	a	a	a	b	a	a	b	a	a	b
	Turkey	b	a	c	a	c	a	e	a	b	a	b	a	d	b	b	a	a	b	a	b	a	a	a	a	a	h	a	a	a	a	a	a
	Hungary, Slovenia	a	a	c	a	a	a	a	a	a	a	a	a	a	a	a	d	a	b	a	a	a	a	a	a	a	h	a	a	a	a	a	a
	Germany	g	a	c	b	a	a	c	a	a	c	a	a	a	a	a	a	a	b	a	a	a	a	a	a	b	g	a	a	a	a	a	a
	Italy, Spain	h	a	c	a	a	a	h	a	a	a	a	a	a	a	a	a	b	b	a	a	a	a	a	a	a	h	a	a	a	a	a	a

Table 4 Character matrix.

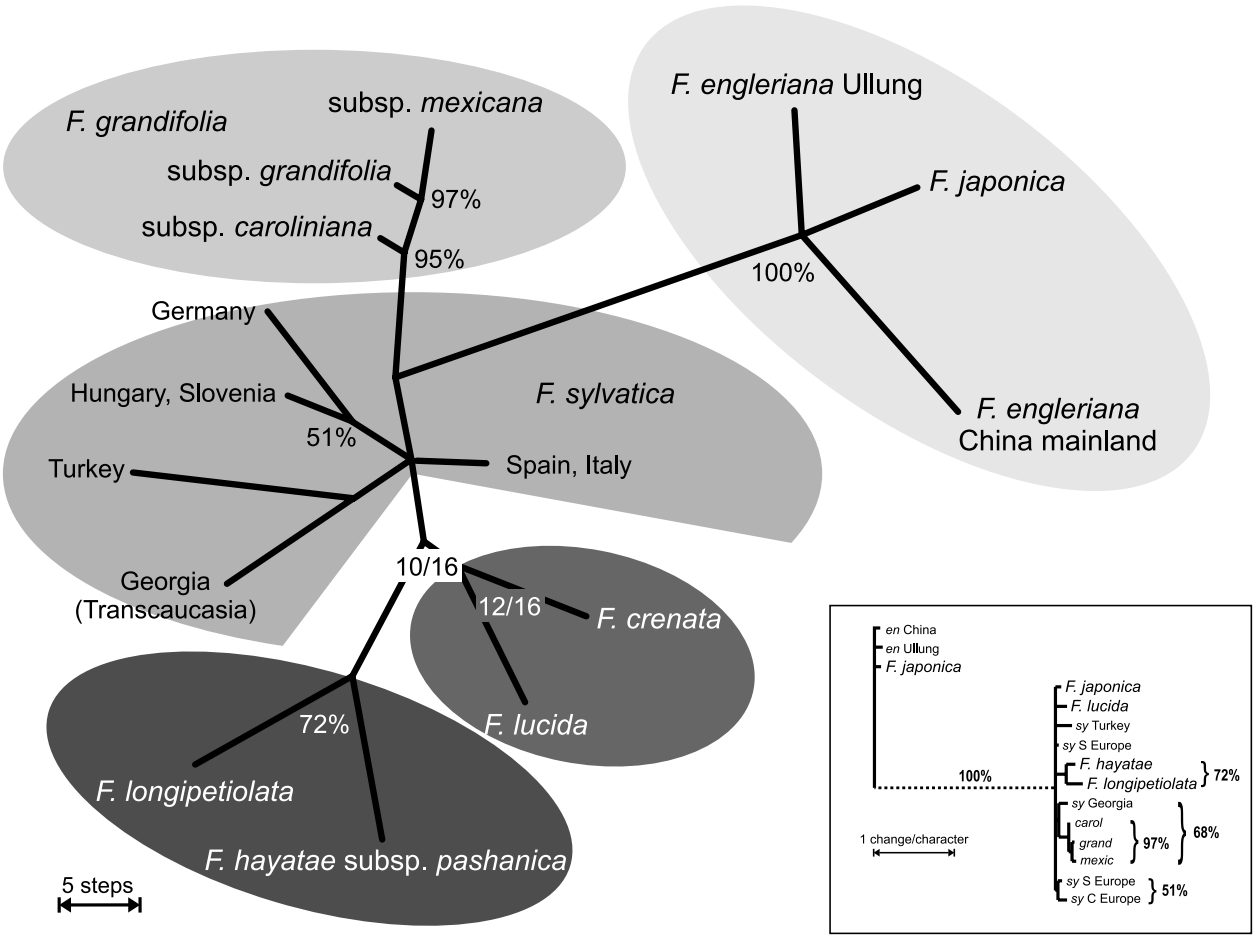

Figure 5 MP phylogram inferred from intraspecific variabilities. The topology shown equals the topology of a majority rule consensus tree of 16 most parsimonious trees (MPT) computed via a branch-and-bound search (PAUP 4.0 beta 10). Branches occurring in less than half of the MPT are ignored. One step is equivalent to the loss or gain of a specific site variability/change of character state. Note that *F. hayatae* subsp. *pashanica* and *F. longipetiolata*, as well as *F. crenata* and *F. lucida* are recognized as sister taxa. Numbers at nodes refer to the number of MPT, which show the according divergence point. Divergence points without numbers occur in all MPT. Percentages at branches indicate posterior probabilities (only >50% shown) computed from the analogously performed Bayesian analysis (inset lower right). Abbreviations: en, *F. engleriana*; carol, grand, mexic: subspecies of *F. grandifolia*; sy, *F. sylvatica.*

into account the different types (binary-unordered, ordered, complex/step-matrices) defined for each character. Secondly, we performed MP and ML runs by setting all characters' types to 'unordered'. By this, the effect of character progression (ordered and complex characters) is minimized and the nucleotide variability of each taxon is then addressed.

Results

To test the hypothesis that intraindividual and intraspecific variability in an ITS data set is reflecting a highly reticulate mode of differentiation and can be used to assess competing tree topologies and to improve phylogenetic resolution, we performed the following analyses for *Fagus*.

Analyses of the matrix using the different character types

A branch-and-bound MP analysis of the data matrix suggests *F. grandifolia* Ehrh. to be nested between the subgenus *Engleriana* and the remaining species of the subgenus *Fagus* (in all 16 most parsimonious trees – MPT; Fig. 5). *Fagus hayatae* subsp. *pashanica* and *F. longipetiolata* are part of a clade comprising *F. crenata*, *F. lucida*, and *F. sylvatica* (all MPT). *Fagus hayatae* subsp. *pashanica* and *F. longipetiolata* are distinguished from the general consensus of *Fagus* and are well supported as sister taxa (all MPT); *Fagus crenata* and *F. lucida* are recognized as sister taxa in most MPT (12/16). Representatives of *F. sylvatica* appear more or less differentiated, but do not form a distinct clade. The ML via BI analysis suggests monophyly of the subgenus *Engleriana* (posterior probability of 100%) and *F. grandifolia* (95%). A closer relationship is further indicated for *F. hayatae* subsp. *pashanica* and *F. longipetiolata* (72%, other alternatives $\leq$ 0.5%) and *F. sylvatica* individuals from Germany, Hungary and Slovenia (51%; in all MPT). Moreover, the Georgian accessions of *F. sylvatica* are placed as a weakly supported sister taxon (68%) to the *F. grandifolia*-clade, which is nested within the subgenus *Fagus* (Fig. 5, inset lower right).

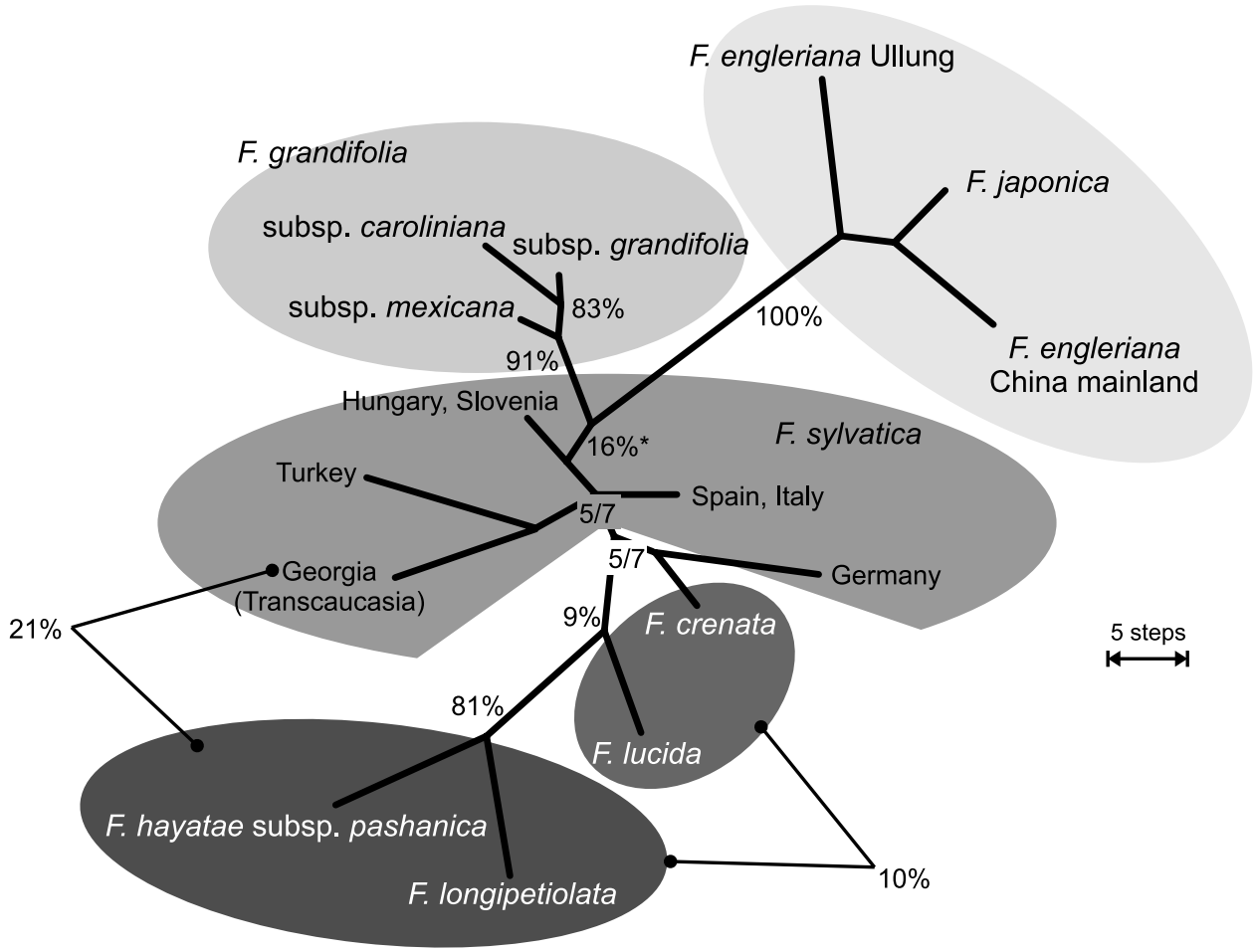

Figure 6 One of seven MPT (most parsimonious trees) computed with character types set to 'unordered', exhibiting a topology identical to the majority rule consensus of all seven MPT. Percentages are Bayesian posterior probabilities for selected nodes. Probabilities for alternative groupings (*F. hayatae/longipetiolata* + Georgian *F. sylvatica*; *F. hayatae/longipetiolata* + *F. crenata* and *F. lucida*) are indicated. * In 24% of the saved ML/BI topologies *F. hayatae* + *F. longipetiolata* are placed as a sister clade to subgenus *Engleriana.*

Analyses with all types of characters set to 'unordered'

Figure 6 shows one of seven MPT with a topology equal to the consensus of all MPT. Again, members of the subgenus *Engleriana* are clearly separated from the taxa of the subgenus *Fagus* (Fig. 5). *Fagus grandifolia* appears as a sister clade to the remaining taxa of subgenus *Fagus*. *Fagus hayatae* subsp. *pashanica* and *F. longipetiolata* consistently come out as sister taxa. In contrast to the first analyses (Fig. 5), all MPT suggest *F. lucida* as sister taxon to *F. hayatae* + *F. longipetiolata.* In addition, *F. crenata* and *F. sylvatica* from Germany form a sister clade to the *F. lucida* + {*F. hayatae* + *F. longipetiolata*} clade in five out of seven MPT. In the remaining two MPT, either *F. crenata* + *F. sylvatica* (Germany) or the *F. lucida-hayatae-longipetiolata* group form a sister clade to the Georgian and Turkish individuals of *F. sylvatica.* The exact position of Georgian and Turkish, southern and eastern European individuals of *F. sylvatica* is not resolved further. Posterior probabilities of the ML via BI analysis support the monophyly of subgenus *Engleriana* (100%), *F. grandifolia* (91%), and the sister relationship between *F. hayatae* and *F. longipetiolata* (81%). Moreover, in 57% of the saved ML/BI topologies, *F. lucida* is placed as sister group to *F. crenata*. It is noteworthy that both analyses (MP and ML) indicate a subsp. *mexicana* + {subsp. *caroliniana* + subsp. *grandifolia*} relationship within *F. grandifolia* (all MPT, sister relationship between subsp. *caroliniana* and *grandifolia* supported at 83%).

Maximum parsimonious reconstruction (MPR) of character evolution

We did not consider absolute substitution probabilities for character coding. Instead, the 'MPRSet' command implemented in PAUP, was used to plot the individual evolution of character states onto a consensus cladogram that best reflects the results of this study and the study by Denk *et al.* (2005). This was done in order to (i) better understand general trends of increase/decrease of intraindividual/intraspecific ITS variability, and (ii) to trace character progression within complex characters (mainly linked site variabilities). In Figures 7 and 8 the increase and loss of genetic variability at the ITS sequence level for some characters is indicated by '+' and '−', respectively. 'Ancestral states', indicated in the middle in Figures 7 and 8, are hypothetical and refer to the results of a multi-evidence study (Denk *et al.*, 2005). The MPR suggests the same trends for binary, ordered, and complex characters both within the

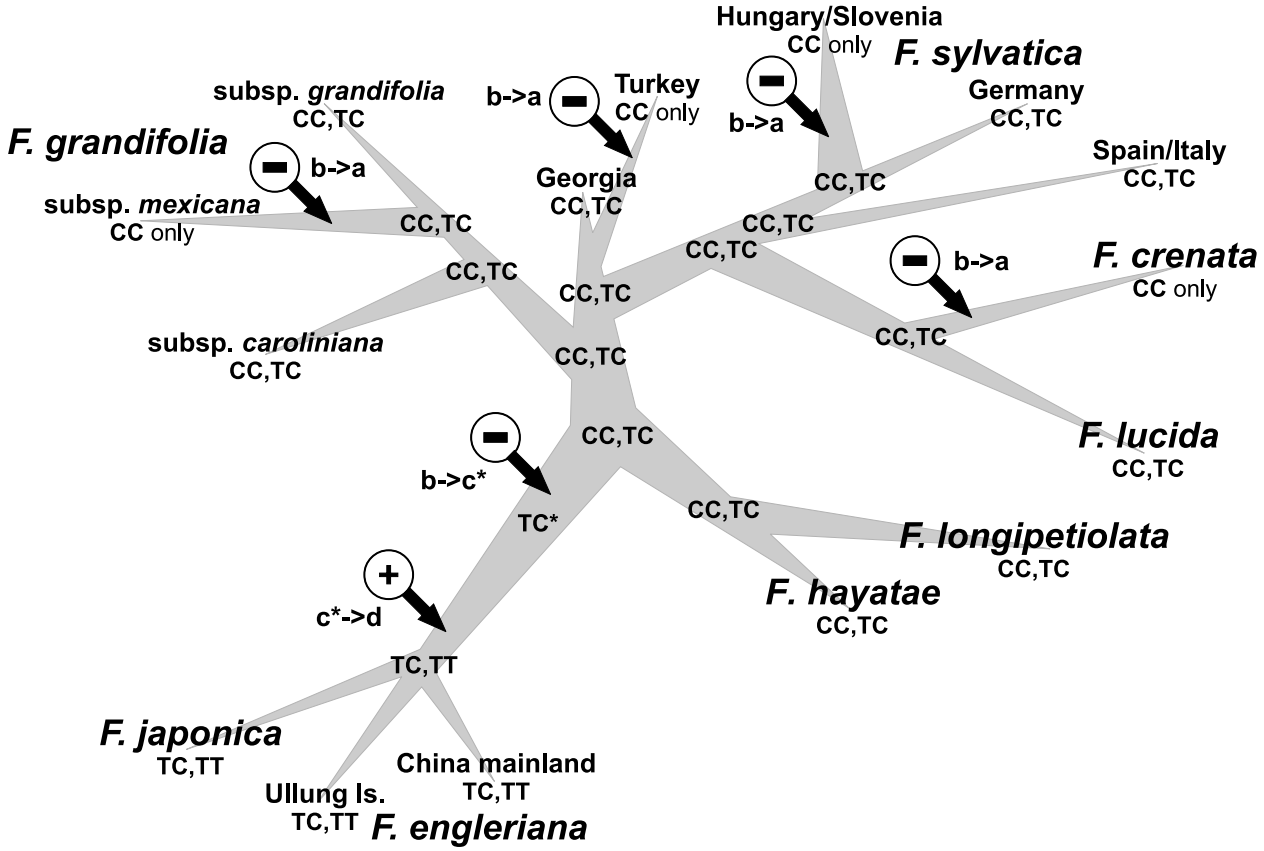

Figure 7 MPR (maximum parsimony reconstruction) for character 1 (2 nucleotides, ordered, 5' ITS1; Table 2). MPR were performed for this and all other characters using the shown cladogram (shaded grey, see text). Reconstructed nucleotide states (CC = 'a' in Tables 3, 4; CC, TC = 'b'; TC = 'c*'; TC, TT = 'd') are indicated for each node. Arrows indicate character state changes (b→a; b→c; c→d). Other symbols: '+', gain of variability; '−', loss of variability; '*', hypothetical character state, not realized in any extant taxon (Tables 2, 4).

ITS1 and ITS2 (details not shown)*: (i) The subgenus *Engleriana* (*F. engleriana, F. japonica*) is characterized by the accumulation of unique (possibly synapomorphic) site variabilities (characters 1, 3; Figs 7, 8); (ii) subspecies of *F. grandifolia* exhibit a progressive transformation from the original variability towards an autapomorphic ITS sequence along a south-north gradient (characters 3, 7; Fig. 8); (iii) the remaining species of the subgenus *Fagus* retain the original ITS variability, new ITS variability is occasionally gained (all Eurasian taxa of the subgenus *Fagus*; characters 3, 9; Fig. 8) and lost in *F. sylvatica* from eastern to south-western and northern populations (characters 1, 3, 9; Figs 7, 8). Comparatively low variability is also detected in *F. crenata* (characters 1, 3, 9; Figs 7, 8).

Discussion

Dependence versus independence of site variabilities and nucleotides

The dependence or independence of distinct characters is difficult to evaluate. The alignment and data in Table 2 suggest dependence of mutations in different parts of the spacer region. Linked (i.e. logically dependent) mutations are not necessarily complementary to each other. In addition, although linked mutations are detected for a certain taxon or group of taxa, the same positions are not necessarily linked for other taxa (Fig. 9; cf. Fig. 4). It is at present not possible to determine whether the observed linkage is due to a change/shift in the secondary structure, general compensatory trends (Torres *et al.*, 1990), occasional parallelism, a general susceptibility for mutation at a specific site, or the formation of co-occurring genotypes as a consequence of incomplete concerted evolution. The secondary structure and folding of the 35S pre-RNA transcript probably have an eminent impact on mutation patterns. Nevertheless, our data indicate that currently available rRNA folding algorithms and software packages are not suitable to fold a hypothetical transcript based solely on the DNA sequence and comprising only a portion of the whole transcript. A secondary structure model for the ITS2 of the Fagales has been developed by Coleman (2003) to correlate ITS2 sequences for higher hierarchical levels (above genus). We detected the highest intraindividual variability in species of the subgenus *Engleriana* (Denk *et al.*, 2005; this study). Putative compensatory base changes can be extracted from the sequence data (Table 5), but are located in the assumed loop and not in the stem regions of the ITS2 structure model presented by Coleman (2003; Table 5). Moreover, base changes are found in the supposedly

* For a coloured illustration including the MPR of all characters see Grimm (2003, Fig. 3–14, p. 49). Minor differences in the reported character states are due to the inclusion of new and additional data for this study.

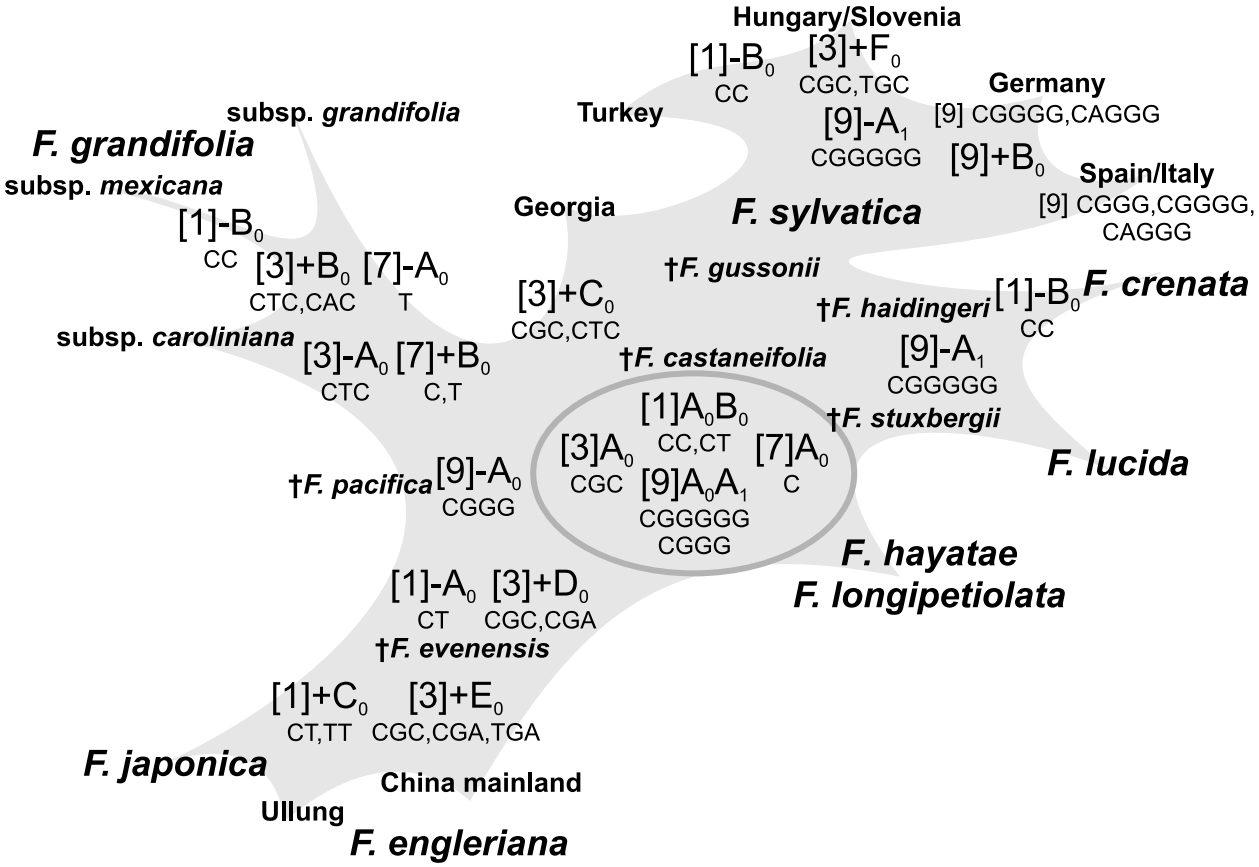

Figure 8 MPR (maximum parsimony reconstruction) of the characters 1 (ordered), 3 (complex), 7 (binary), 9 (complex) mapped onto the proposed differentiation history of the genus *Fagus* (shaded grey; based on Denk *et al.*, 2005). Note that fewer character changes are required when a dynamic evolutionary framework is applied instead of the static cladogram used for the MPR (Fig. 7; see discussion). Reconstructed and actual nucleotide composition is indicated for each character change. Putative 'ancestral states' are encircled.

structurally conserved stem regions that are not compensated. Apparently, new rRNA folding algorithms have to be invented to develop reliable structure models for the rDNA spacer sequences. The unreliability of the most commonly used folding algorithm and software (*Mfold 3.1*; Zuker, 2003) has recently been demonstrated for the 16S/18S and 23S/28S rRNA by Doshi *et al.* (2004). Hence, its application to hypothetical transcripts of non-coding rDNA spacers is even more critical.

ITS evolutionary processes do not appear to influence each single nucleotide individually and independently. Instead, the complete sequences or a number of partial sequences of the ITS1 and ITS2 have to be considered as one/few evolving character complex/complexes. For the tree genus *Zelkova*, Ulmaceae, Denk and Grimm (2005) defined series of mutations and length polymorphic regions as distinctive oligonucleotide motives and reconstructed pathways of molecular evolution for each motif. The deduced evolutionary pathway of motive variants was found to follow a generally parsimonious pattern and was in accordance with the substitution probability for each assumed mutation event. The distribution of different variants of oligonucleotide motives in the ITS of morphologically distinct taxa suggested relatively recent horizontal gene flow between species of *Zelkova*. Adding evidence from the fossil record speciation processes and migration routes could be reconstructed for the Late Cenozoic.

Intraindividual and intraspecific ITS variability in *Fagus* is markedly higher (up to 0.1 divergence in *F. engleriana/F. japonica*; Denk *et al.*, 2005) and, in addition, consensual and putatively 'derived' sequence motives (cf. Denk *et al.*, 2005) and nucleotide states are strongly intermixing within the same ITS1 and ITS2 variant (cf. Table 5). Thus, a detailed reconstruction of the molecular evolution of the complete ITS1 and ITS2, as performed for *Zelkova*, has to remain a task for the future.

Linked nucleotides and character complexes (cf. Denk *et al.*, 2005) appear to support the assumption that the detected ITS variability is in fact originating from paraloguous variants. It cannot be ruled out that the ITS sequences assembled represent more or less independently evolving lineages (homoeologues or 'paralogues' in a broad sense); the multiple loci of the rRNA gene obviously are susceptible to retain paraloguous data (Àlvarez & Wendel, 2003; Volkov *et al.*, 2004; cf. Denk & Grimm, 2005, for *Zelkova*, and Denk *et al.*, 2005, for *Fagus*). Moreover, the number of nucleolus organisation regions and chromosome numbers are not available for all *Fagus* species and have never been studied at the population level. ITS variants can be distinguished to a certain degree in *Fagus*, in particular in case of the highly divergent and apparently derived subgenus *Engleriana* (Table 5; cf. Denk *et al.*, 2005). For example, clones exhibiting the

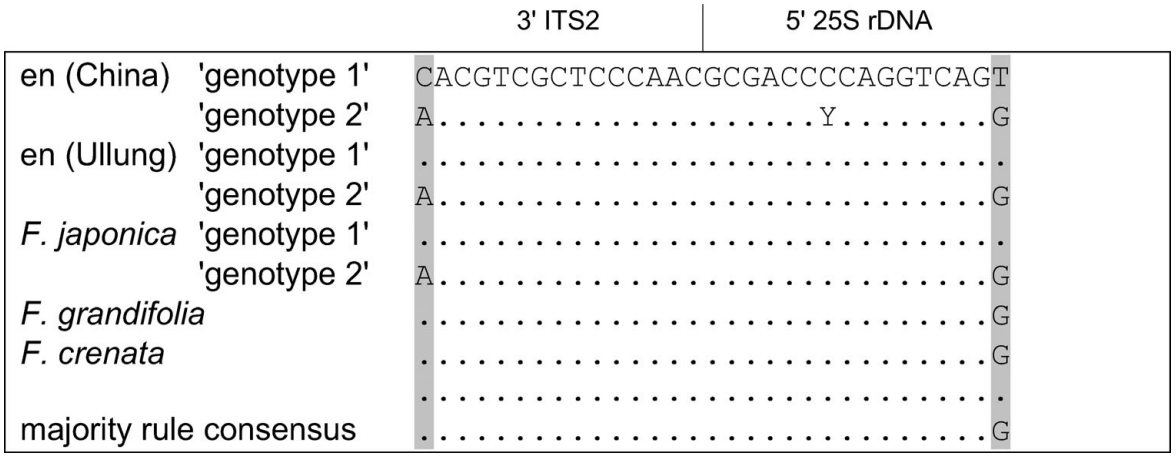

Figure 9 3' end of ITS2 and 5' end of 25S rDNA showing a conspicuous dependence between two site variabilities in certain *Fagus* spp. (grey background). Most *Fagus* spp. do not show any site variability at all, instead a C at the first position is always correlated to a G at the last position shown (i.e. majority rule consensus). All accessions of the subgenus *Engleriana* (a total of 28 clones) can be assigned either to a genotype 1 (C with T) or genotype 2 (A with G), both differing from the above-mentioned consensus. Accessions of *F. crenata* are either identical to the consensus or to genotype 1. Abbreviations: 'en', *F. engleriana*. Dots indicate nucleotides identical to the top sequence.

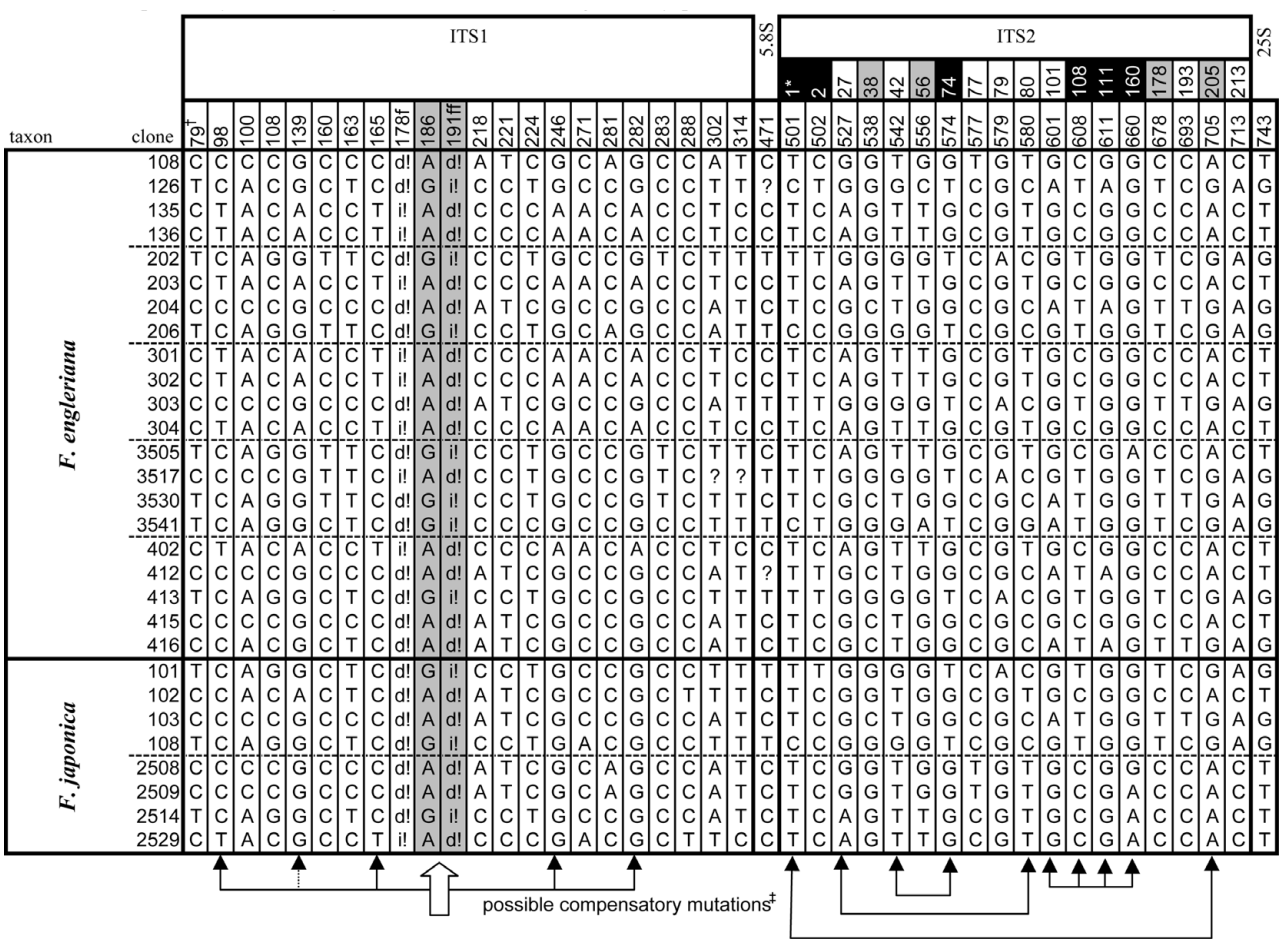

Table 5 Compensatory base changes in ITS variants of *F. engleriana/japonica*.

*Site number according Coleman's (2003) secondary structure model: not shaded, located within loop/unpaired stem; grey, located within stem; black, highly conserved site in respect to all Fagales (not necessarily paired).

†Site number of herein used alignment.

‡As inferred from the nucleotide structure of ITS clones, not in accordance with the predicted secondary structure model. Thick arrow indicates a linkage between the 13bp long *Engleriana*-indel and an upstream nucleotide.

prominent autapomorphic ITS1 insertion (pos. 191ff, Table 5; putative 'homoeologue 1': en-126, -202, -206, -413, -3505, -3530, -3541, ja-101, -108, -2514) are similar in the ITS1 (e.g. pos. 79, 98, 100 always T-C-A in contrast to C-T-A, C-C-C in other clones: 'homoeologue 2'). However, while the ITS2 of 'homoeologue 1' sequences can be markedly different even within the same individual, identical ITS2 sequences can be found within 'homoeologue 1' and 'homoeologue 2' sequences originating from different and geographically isolated individuals (e.g. en-202, ja-101 and en-413, -3517). Therefore, the detected ITS variants (each clone represents one tandem repeat) do not form discrete groups of homoeologues or 'ITS paralogues' as had been assumed for other genera (Àlvarez & Wendel, 2003; Bailey *et al.*, 2003, and references cited there), but are apparently subject to frequent intragenomic recombination.

Phylogenetic and systematic implications

The results of the present study are in accordance with a most recent morphological (Denk, 2003) and a multi-evidence study (Denk *et al.*, 2005). The subgenus *Engleriana* appears to be morphologically and genetically strongly derived in relation to extant taxa of the subgenus *Fagus* and fossil representatives of the genus (Figs 5–8; cf. Denk *et al.*, 2005). Here, the potentially derived state of the ITS is indicated by numerous point mutations that cannot be detected in the ITS of any other *Fagus* species and, accordingly, by conspicuous site variabilities confined to *F. engleriana* and *F. japonica* (Table 2; Fig. 4); but these site variabilities generally are characterized by the co-occurrence of a consensual (possibly 'ancestral' or symplesiomorphic) nucleotide state with one or more putatively synapomorphic nucleotide state(s).

Our analyses suggest that *F. grandifolia* is nested between the subgenus *Engleriana* and the remaining taxa of the subgenus *Fagus* (MPT of both runs). However, Bayesian posterior probabilities are below <50%. A possible indication for a sister relationship between *F. grandifolia* and the subgenus *Engleriana* may be the northern Pacific origin of the genus (Manchester & Dillhoff, 2004; Denk *et al.*, 2005). In fact, details of the MPR (see e.g. character 9 in Fig. 8) and the data presented in Denk *et al.* (2005) corroborate the assumption that morphological and ITS features shared between subgenus *Engleriana* and *F. grandifolia* are symplesiomorphic in a strict cladistic sense and predate the split of the subgenus *Engleriana* from the subgenus *Fagus* and the speciation of *F. grandifolia* as the only lineage east of the Pacific.

A sister relationship between *F. hayatae* and *F. longipetiolata* has previously been suggested by Denk (2003) and Denk *et al.* (2005) and is statistically confirmed by coding intraindividual/-specific ITS polymorphisms as characters (Figs 5, 6). The MPR of these characters (Figs 7, 8) further support the interpretation of *F. hayatae* and *F. longipetiolata* as 'genetic living fossils' from an ITS viewpoint as already proposed by Denk *et al.* (2005).

A closer relationship between *F. crenata* and *F. lucida* as indicated by the first MP analysis (using characters' predefined types) and the second ML analysis (ignoring characters' types) is in conflict with morphological evidence (but see Shen, 1992), which points towards a closer relationship between *F. sylvatica* and *F. crenata.* Denk *et al.* (2005) found that ITS oligonucleotide motives in *F. lucida* comprise motif variants that are otherwise typical of *F. hayatae* and/or *F. longipetiolata*, but are never found in the ITS of *F. crenata* and *F. sylvatica*. Hence, these variants support the topology shown in Figure 6. One reason for this could be that gene-flow occurred between a lineage ancestral to *F. lucida* and ancestors of *F. crenata/sylvatica* and *F. hayatae/longipetiolata.* The latter possibility is favoured by the recent geographic distribution of the modern species.

Figures 4, 7 and 8 show that molecular differentiation in the ITS of *Fagus* is predominately related to the gain and loss of intraindividual variability. Even the most derived taxa *F. engleriana* and *F. japonica* (cf. Denk *et al.*, 2005) show some of the consensual nucleotide states, which allows us to hypothesize about the ancestry of ITS characteristics. This is exemplarily illustrated in Figure 10 with the complex character 3 (Table 2). Intraindividual variability may have two reasons: (i) intrapopulation level differentiation in combination with incomplete concerted evolution, and (ii) horizontal gene flow across population and species boundaries. Possibly incomplete concerted evolution can be observed in *F. grandifolia*, where a more consensual and 'ancestral' ITS (subsp. *mexicana*) subsequently is lost in individuals of subsp. *caroliniana* and *grandifolia* (south-eastern USA, north-eastern USA). This may partly apply also to *F. engleriana/japonica,* because all detected ITS variants combine 'derived' and assumedly 'ancestral' nucleotide states. Retention and increase of intraindividual variability through time (genetic living fossils *F. hayatae, F. longipetiolata*) is all the more likely when the population dynamics of *Fagus* and its pronouncedly stenoecious behaviour are taken into account. The strongly overlapping and distinctive variability patterns found in morphological and ITS characters of particular groups (e.g. *F. hayatae, F. longipetiolata* and *F. lucida*, subgenus *Engleriana* and *F. grandifolia* and *F. longipetiolata*; Eurasian taxa of the subgenus *Fagus*) can only be explained by long phases of unhindered horizontal gene flow (at least for the ITS locus), i.e. strongly reticulate evolution. This is in line with the absence of pronounced morphological boundaries in early northern hemispheric fossils of *Fagus*, and between the Late Cenozoic ancestors of morphologically distinct modern species. Morphological and molecular (ITS) patterns observed across a wide geographical range of *F. sylvatica* (Denk *et al.*, 2002) may provide a clue to understanding early population and species differentiation processes in *Fagus.*

Conclusions

Intraspecific and intraindividual variability are likely to disguise phylogenetic relationships when standard analytical methods are employed, especially if phylogenies are based on a rather variable gene region such as the ITS that is inherited by both parental lineages (Álvarez & Wendel, 2003). The reduction of molecular evolution to simple mutational categories (A↔C, A↔G, etc.) appears not to be sufficient to fully

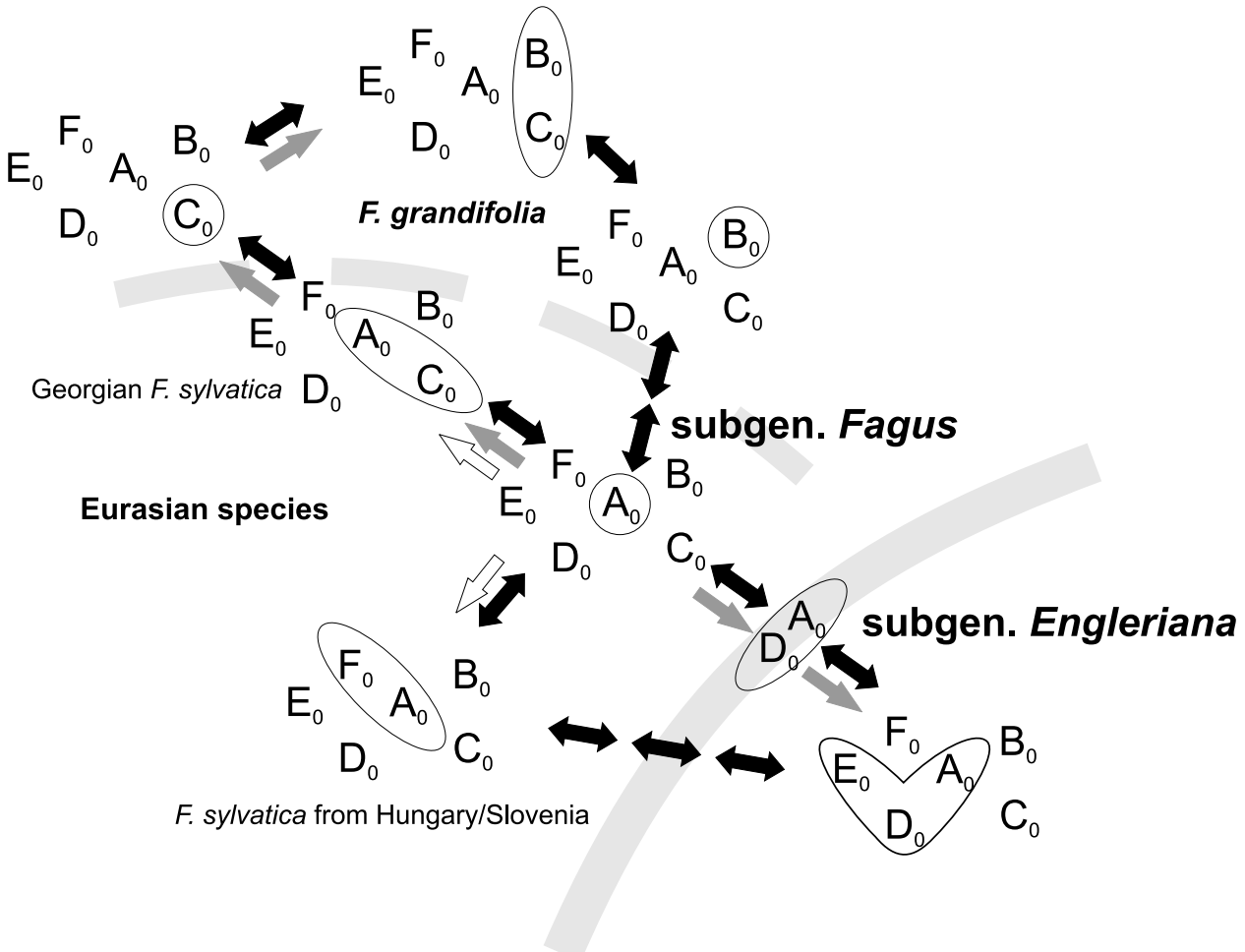

Figure 10 Differentiation of ITS gene pools for character 3 according to the step-matrix coding (black arrows, each arrow equals one step) and MPR (maximum parsimonious reconstruction; grey arrows). White arrows indicate differentiation into subpopulations of *F. sylvatica.* Strong correlation between the coding of complex characters and the assumed phylogeny (cf. Fig. 7) is obvious. Note that *F. grandifolia*, subgenus *Engleriana*, and all extinct Eurasian taxa of the subgenus *Fagus* can be easily derived from a heterogeneous and polymorphic ancestral gene pool ($\{A_0C_0\} + A_0$; cf. Fig. 8.

recognize phylogenetic information provided by gaps, and, in case of reticulate evolution, by site variabilities. We are aware that the coding procedure presented here cannot generally replace statistical 'base-to-base' analysis methods but we believe that it does help analysing data sets, in which intraspecific and intraindividual variabilities are as high or almost as high as the overall interspecific variability. Such is the case for the ITS of *Fagus* that was used as model system in the present study. Comparatively low overall genetic differentiation combined with high intraspecific variability resulted in a data set that could not be fully resolved with 'base-to-base' analyses. Using the information provided by intraindividual and intraspecific variability helps to resolve further intrageneric differentiation within *Fagus* (Denk *et al.*, 2005; this study), e.g. concerning the character progression in *F. grandifolia* along the putative migration route during the Cenozoic (Northern Pacific basin [fossil] – Mexico ['Tertiary' relict] – eastern North America), or the systematic position of *F. lucida*.

Studies dealing with subgeneric relationships in widespread tree genera such as *Fagus* may be strongly affected by recent and fossil hybridization events and/or incomplete concerted evolution (Dover & Tautz, 1986). Extant plant species possibly originated from complex reticulate evolution and have a complex biogeographic, migration and speciation history. Thus, we believe that comprehensive assembling and use of genetic variability of closely related plant taxa is crucial for reconstructing a sound phylogeny based on molecular markers. Combined with data from other sources such as biogeography, ecology, morphology and the fossil record, it should be possible to achieve more reliable and precise reconstructions of low-level evolution, and, eventually, a better understanding of the molecular differentiation in the course of speciation processes.

Acknowledgements

Special thanks are due to Karin Stögerer for her skilful technical assistance in assembling the data. T. Denk acknowledges funding of a field trip to China by the Swedish Research Council. Laboratory work was funded by the German Science Foundation (DFG).

References

Ackerly, D.D. & Donoghue, M.J. 1998. Leaf size, sapling allometry, and Corner's rules: a phylogenetic study of correlated evolution in maples (*Acer*). *American Naturalist* **152**, 767–791.

Álvarez, I. & Wendel, J.F. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**, 417–434.

Bailey, C.D., Carr, T.G., Harris, S.A. & Hughes, C.E. 2003. Characterization of angiosperm nrDNA polymorphism, paralogy, and pseudogenes. *Molecular Phylogenetics and Evolution* **29**, 435–455.

Baldwin, B.G., Sanderson, M.J., Porter, J.M., Wojciechowski, M.F., Campbell, C.S. & Donoghue, M.J. 1995. The ITS region of nuclear ribosomal DNA: A valuable source of evidence of angiosperm phylogeny. *Annals of the Missouri Botanic Gardens* **82**, 247–277.

Coleman, A.W. 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparison. *Trends in Genetics* **19**, 370–375.

Denk, T. 2003. Phylogeny of *Fagus* L. (Fagaceae) based on morphological data. *Plant Systematics and Evolution* **240**, 55–81.

Denk, T. & Grimm, G.W. 2005. Phylogeny and biogeography of *Zelkova* (Ulmaceae *sensu stricto*) as inferred from leaf morphology, ITS sequence data and the fossil record. *Botanical Journal of the Linnean Society* **147**, 129–157.

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.

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.

Doshi, K.J., Cannone, J.J., Cobaugh, C.W. & Gutell, R.R. 2004. Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction. *BMC Bioinformatics* **5**, 105.

Dover, G.A. & Tautz, D. 1986. Conservation and divergence in multigene families: Alternatives to selection and drift. *Philosophical Transactions of the Royal Society London Series B* **312**, 275–289.

Forey, P.L., Humphries, C.J., Kitching, I.J., Scotland, R.W., Siebert, D.J. & Williams, D.M. 1992. *Cladistics, a Practical Course in Systematics*. Clarendon Press, Oxford.

Grimm, G.W. 2003. Tracing the mode and speed of intrageneric evolution – a phylogenetic case study on genus *Acer* L. (Aceraceae) and *Fagus* L. (Fagaceae) using fossil, morphological, and molecular data. D. Sc. thesis, Tübingen, 159 pp + appendices. URN: urn:nbn:de:bsz:21-opus-15744

Huelsenbeck, J.P. & Ronquist, F. 2001. MrBayes: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755.

Jobst, J., King, K. & Hemleben, V. 1998. Molecular evolution of the internal transcribed spacers (ITS1 and ITS2) and phylogenetic relationships among species of Cucurbitaceae. *Molecular Phylogenetics and Evolution* **9**, 204–219.

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.

Manos, P.S. & Stanford, A.M. 2001. The historical biogeography of Fagaceae: Tracking the Tertiary history of temperate and subtropical forests of the Northern Hemisphere. *International Journal of Plant Science* **162**, S77–S93.

Shen, C.-F. 1992. A monograph of the genus *Fagus* Tourn. ex. L. (Fagaceae). Ph.D. thesis, The City University of New York, 390 pp.

Stanford, A.M. 1998. The biogeography and phylogeny of *Castanea*, *Fagus*, and *Juglans* based on *mat*K and ITS sequence data. Ph.D. thesis, Chapel Hill, 261 pp.

Suh, Y., Heo, K. & Park, C.-W. 2000. Phylogenetic relationships of maples (*Acer* L.; Aceraceae) implied by nuclear ribosomal ITS sequences. *Journal of Plant Research* **113**, 193–202.

Swofford, D.L. 2002. PAUP*: Phylogenetic analysis using parsimony (* and other methods), version 4, computer program distributed through the Illinois Natural History Survey, Champaign, USA.

Tian, X., Guo, Z.-H. & Li, D.-Z. 2002. Phylogeny of Aceraceae based on ITS and *trn*L-F data sets. *Acta Botanica Sinica* **44**, 714–724.

Torres, R.A., Ganal, M. & Hemleben, V. 1990. GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal DNA. *Journal of Molecular Evolution* **30**, 170–181.

Volkov, R.A., Medina, F.J., Zentgraf, U. & Hemleben, V. 2004. Molecular cell biology: Organization and molecular evolution of rDNA, nucleolar dominance and nucleolus structure. In: Esser, K., Lüttge, U., Beyschlag, W. & Murata, J., Eds., *Progress in Botany 65*. Springer Verlag, Berlin, pp. 106–146.

Wheeler, W. 1999. Fixed character states and the optimisation of molecular sequence data. *Cladistics* **15**, 379–385.

Wheeler, W. 2001. Homology and the optimisation of DNA sequence data. *Cladistics* **17**, S3–S11.

Wissemann, V. 2003. Hybridization and the evolution of the nrITS spacer region. In: Sharma, A.K. & Sharma, A., Eds., *Plant Genome: Biodiversity and Evolution 1A*. Science Publishers Inc., Enfield, Plymouth, pp. 57–71.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acid Research* **31**, 3406–3415.

accession numbers	clone numbers	assigned taxon	origin
AY232915-AY232918	cr-2xx	*F. crenata*	Honshu, Japan
AF456967-AF456970, AF456975, AF457014	cr-30xx		Botanical Garden, Univ. of Tübingen, seed imported from Japan
AY232893-AY232903, AY232986, AY232991	en-1xx, en-2xx, en-3xx	*F. engleriana*	3 localities, Hubei province, China
AY232905-AY232908, AY232987, AY232992	en-4xx		Seo-Myun, western part of Ullung Do, S Korea
AF456981, AF456982, AF457020, AF457021, AY232904	en-35xx		Botanical Garden, Univ. of Tübingen, seed imported from China
AY232922-AY232926	gr-2xx, gr-6xx	*F. grandifolia* ssp. *caroliniana*	Atlanta/Florida Caverns S.P., S Georgia/N Florida, U.S.
AF456976-AF456978, AF456980, AF457015-AF457017, AF457019, AY232919-AY232921	gr-26xx, gr-27xx	*F. grandifolia* ssp. *grandifolia*	2 localities, New York State, U.S.
AY232927-AY232930	gr-51xx	*F. grandifolia* ssp. *mexicana*	Zacualtipán, Hidalgo, Mexico
AY232931-AY232942, AY232988, AY232993	ha-3xx, ha-4xx, ha-5xx	*F. hayatae* ssp. *pashanica*	3 localities, Hubei province, China
AF456971, AF456972, AY232909-AY232914	ja-1xx, ja-25xx	*F. japonica*	2 localities, Honshu, Japan
AY232943-AY232954	lo-1xx, lo-2xx, lo-3xx	*F. longipetiolata*	3 morphotypes, Hubei province, China
AF456973, AF456973, AF457012, AF457013, AY232955, AY232956	lo-47xx		Fujian province, China
AY232957-AY232960	lu-1xx	*F. lucida*	SW Hubei, near border to Sechuan, China
AY232961-AY232963, (AF456926)	lu-48xx		Guizhou province, China
AF456983-AF456989	ho-16xx, ho-18xx, ho-19xx	*F. sylvatica*	3 localities, Georgia, Transcaucasia
AF456938-AF456945, AF456951-AF456953, AY232964-AY232968, AY232989, AY232994	or-4xx, or-6xx, or-12xx or-13xx		4 localities, N Turkey
AY232969-AY232971, AY232990, AY232995	sy-20xx		János-hegy, Hungary
AF457008-AF457011, AF457047-AF457050	sy-43xx		Podcetrtek, Slovenia
AF456995-AF457003, AF457034-AF457042	sy-28xx, sy-29xx, sy-31xx sy-32xx		2 localities, C and S Germany
AY232978-AY232985	sy-46xx, sy-47xx, sy-48xx, sy-49xx		3 localities, N Italy
AF456993, AF456994, AF457032, AF457033, AY232972-AY232977	sy-16xx, sy-54xx, sy-55xx		3 localities, N Spain

Appendix: Accession numbers and origin of sampled individuals.