UNIVERSITÀ
DEGLI STUDI
FIRENZE

FLORE
Repository istituzionale dell'Università degli Studi di Firenze

Utility of ITS Sequence Data for Phylogenetic Reconstruction of Italian Quercus spp.

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

Original Citation:

Utility of ITS Sequence Data for Phylogenetic Reconstruction of Italian Quercus spp / R. Bellarosa; M. Simeone; A. Papini; B. Schirone. - In: MOLECULAR PHYLOGENETICS AND EVOLUTION. - ISSN 1055-7903. - STAMPA. - 34:(2005), pp. 355-370.

Availability:

This version is available at: 2158/341836 since:

Terms of use:

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf)

Publisher copyright claim:

(Article begins on next page)

23 April 2024

Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Molecular Phylogenetics and Evolution 34 (2005) 355–370
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Utility of ITS sequence data for phylogenetic reconstruction of Italian *Quercus* spp.

Rosanna Bellarosa[a,*], Marco C. Simeone[a], Alessio Papini[b], Bartolomeo Schirone[a]

[a] *Dipartimento di tecnologie, ingegneria e scienze dell'Ambiente e delle Foreste (D.A.F.), Università degli Studi della Tuscia, via S. Camillo de Lellis, 01100 Viterbo, Italy*

[b] *Dipartimento di Biologia Vegetale, Università degli Studi di Firenze, via La Pira 4, 50121 Firenze, Italy*

Received 25 June 2003; revised 24 September 2004

Abstract

Nuclear ribosomal DNA sequences encoding the 5.8S RNA and the flanking internal transcribed spacers (ITS1 and ITS2) were used to test the phylogenetic relationships within 12 Italian *Quercus* taxa (Fagaceae). Hypotheses of sequence orthology are tested by detailed inspection of some basic features of oak ITS sequences (i.e., general patterns of conserved domains, thermodynamic stability and predicted conformation of the secondary structure of transcripts) that also allowed more accurate sequence alignment. Analysis of ITS variation supported three monophyletic groups, corresponding to subg. *Cerris*, *Schlerophyllodrys* (= *Ilex* group *sensu* Nixon) and *Quercus*, as proposed by Schwarz [Feddes Rep., Sonderbeih. D, 1-200]. A derivation of the "*Cerris* group" from the "*Ilex* group" is suggested, with *Q. cerris* sister to the rest of the "*Cerris* group." *Quercus pubescens* was found to be sister to the rest of the "*Quercus* group." The status of hybrispecies of *Q. crenata* (*Q. cerris* × *Q. suber*) and *Q. morisii* (*Q. ilex* × *Q. suber*) was evaluated and discussed. Finally, the phylogenetic position of the Italian species in a broader context of the genus is presented. The utility of the ITS marker to assess the molecular systematics of oaks is therefore confirmed. The importance of Italy as a region with a high degree of diversity at the population and genetic level is discussed.
© 2004 Elsevier Inc. All rights reserved.

Keywords: Phylogeny; *Quercus*; ITS sequences; Secondary structures

1. Introduction

Italy has a high diversity of native oaks, with up to 15 species and a series of hybrids and ecotypes (Pignatti, 1982). This diversity is outstanding considering the fact that there are only 17 fully recognised *Quercus* species in Europe (including Turkey) as well as a group of about 10 entities with an uncertain taxonomic rank (Schwarz, 1964). Moreover, oak species with Eastern (e.g., *Q. trojana*, *Q. macrolepis*), Western (*Q. suber*), Central (*Q. petraea*, *Q. robur*) European distributional range and with evergreen (*Q. ilex*, *Q. coccifera*, *Q. suber*), semi-deciduous (*Q. trojana*, *Q. frainetto*), and deciduous habits (*Q. cerris*, *Q. pubescens*, *Q. robur*, *Q. petraea*) can be found in the Italian woodlands. This ecological and species richness can be explained by the peculiar geographical position of the Italian peninsula and its role as a glacial refugium, e.g., an area with a milder climate where the temperate forest species were confined (and protected from starvation) for most of the Glacial Period, before the recolonisation of Central and Northern parts of Europe could take place (Bennet et al., 1991; Brewer et al., 2002; Hewitt, 1996; Schirone and Spada, 1995; Taberlet et al., 1998). As generally acknowledged (see for instance Petit et al., 2002a), European glacial refugia, located in the three Mediterranean peninsulas (Italy, Spain, and Greece), are characterised by high degrees of genetic diversity at the population level and by the local persistence of very ancient genotypes with atavistic

* Corresponding author. Fax: +39 0761 357413.
E-mail address: bellaros@unitus.it (R. Bellarosa).

1055-7903/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2004.10.014

characters (Schirone and Spada, 2001). All this makes Italy an ideal laboratory to study the autoecology of single oak species and indeed it offers many stimulating opportunities to investigate the interspecific relationships within the genus *Quercus* from the ecological, genetical, and evolutionary point of view.

Quercus comprises 350–500 species distributed throughout the Northern Hemisphere (Kubitzki, 1993; Nixon, 1993). They are conspicuous members of temperate deciduous forests of North America, Europe, Asia, as well as the evergreen Mediterranean maquis. Several taxonomic treatments have been produced for *Quercus* based on morphology (e.g., Camus, 1936–1954; Nixon, 1993; Schwarz, 1936–1939, 1964; Trelease, 1924). These treatments also show the taxonomic controversies mainly due to the intraspecific morphological variations that may be produced by hybridisation (Burger, 1975) and adaptation to ecological changes in the environment (Van Valen, 1976). As a result, the classification at the subgeneric, sectional, and subsectional levels is still uncertain, and so is the taxonomic ranking of some taxa.

Molecular markers (e.g., storage proteins, allozymes, IGS restriction maps, and cpDNA polymorphism) were used to construct phylogenetic relationships (cf. Bellarosa et al., 1990, 1996; Toumi and Lumaret, 2001), and define the geographic structure of genetic diversity and post-glacial recolonisation routes (Ferris et al., 1993; Fineschi et al., 2002; Petit et al., 2002b, 1997) in oaks. Recently, two research teams (Manos et al., 1999; Samuel et al., 1998) attempted to assess the molecular systematics of *Quercus* using independently generated sequence data of the internal transcribed spacer (ITS) regions. Their reconstructions, however, showed disagreements in aspects, although the phylogenetic trees for both hypotheses were strongly supported by bootstrap resampling. In particular, the maximum parsimony tree derived by Samuel et al. (1998) conflicted with available infrageneric *Quercus* classifications (mostly those concerned with sections *Cerris* and *Quercus*) and documented polyphyletic lines for sect. *Cerris*, thus suggesting Tertiary transcontinental connections between eastern Asia and North America. In fact, in the first clade the deciduous North American *Q. rubra* (sect. *Lobatae*) grouped together with more distant Eurasian species belonging to sect. *Cerris*, with the evergreen West Mediterranean *Q. suber* at the basal position. In the second clade, the deciduous North American *Q. virginiana* (sect. *Quercus*) was basal to the remaining tree branches grouping together with evergreen and deciduous Eurasian species of sections *Cerris* and *Quercus*. Conversely, Manos et al. (1999) provided gene trees with clades that corresponded to well defined taxonomic groups and supported the recognition of the monophyly of the strictly Eurasian section *Cerris*. Mayol and Rossellò (2001) compared the two ITS data sets and detected high levels of variation in the G + C content, as well as large indels responsible for length variation, high rates of substitution in the conserved motifs and differing secondary structures stability. The convincing explanation proposed for the contrasting phylogenetic history derived by Samuel et al. (1998) was that it was generated by a spurious analysis, where some species were represented by ITS pseudogene sequences and only a few by functional orthologues. More recently, the potential risk of mixing data from the contemporary analysis of unnoticed divergent paralogues was also underlined by the study of Muir et al. (2001), who detected three divergent rDNA clusters in *Q. robur* and *Q. petraea*, based on 5.8S and ITS2 sequences. Remarkably, despite the high number of sequences analysed for each gene family, of which only one proved to be functional, the two oak species could not be differentiated.

In this context (intra-individual nrDNA polymorphism, a network of paralogous sequences available in GenBank, difficulty in discriminating closely related taxa), the molecular data by Manos and co-workers (1999, 2001) provided the most reliable phylogenetic reconstruction of *Quercus*. However, they presented limited data from European taxa, which have been traditionally included in various morphologically distinct groups. The focus of this study was, therefore, to test the utility of ITS sequences as an informative tool to draw phylogenetic inferences within Italian *Quercus* species. Specifically, we pursued the objectives of (1) testing the previously proposed taxonomic and phylogenetic schemes with data from ITS sequence variation; (2) providing information from taxa not included in previous phylogenetic reconstructions, for a broader analysis of *Quercus* evolution; and (3) establishing a reliable phylogenetic framework for the main Italian oaks, so that controversial taxa and remaining European species will be included in future studies. Our investigations mainly dealt with sections *Cerris* s.l. (Camus, 1936–1954; subg. *Cerris* plus subg. *Ilex* Nixon, 1993; Schwarz, 1936–1939, 1964) and *Quercus*, to which all Italian oaks belong, focusing on the 10 taxa which are the most widespread and best circumstantiated on a morphological, biological and ecological basis as well as on the two most easily recognisable oak hybrids (*Q. crenata* and *Q. morisii*). A strategy of detailed inspection of some basic features of the sequences (e.g., evaluation of patterns of nucleotide substitution, estimates of stability and uniformity of secondary structures of transcripts) was undertaken to corroborate hypotheses of orthology.

2. Materials and methods

2.1. Sampling, DNA extraction, ITS amplification, and analysis

Plant material was collected from natural Italian populations (Table 1; *exsiccata* available by the authors).

Table 1
The 12 species (16 samples) of the genus *Quercus* surveyed in the present study

Taxa	Source	GenBank Accession Nos.
Subgenus *Cerris*		
Q. cerris L. (1)	C-W Italy (Latium, Cimini mts.)	AY226832
Q. cerris L. (2)	C-E Italy (Majella Natl. Park)	AY226833
Q. crenata Lam.	C-W Italy (Latium, Cimini mts.)	AY226844
Q. macrolepis Kotschky	S-E Italy (Salento Peninsula)	AY226845
Q. morisii Borzì	S-E Italy (Salento Peninsula)	
Q. suber L. (1)	C-W Italy (Thyrrenian coast)	AY226841
Q. suber L. (2)	N Sardinia (Sassari)	AY226842
Q. trojana Webb.	S-E Italy (Salento Peninsula)	AY226843
Subgenus *Quercus*		
Q. frainetto Ten.	S-E Italy (Salento Peninsula)	AY226835
Q. petraea Liebl.	C-W Italy (Latium, Cimini mts.)	AY226838
Q. pubescens Willd. (1)	C-W Italy (Latium, Cimini mts.)	AY226839
Q. pubescens Willd. (2)	C-E Italy (Majella Natl. Park)	AY226846
Q. robur L.	C-W Italy (Latium, Cimini mts.)	AY226846
Subgenus *Schlerophyllodrys*		
Q. coccifera L.	S-E Italy (Salento Peninsula)	AY226834
Q. ilex L. (1)	C-W Italy (Thyrrenian Coast)	AY226836
Q. ilex L. (2)	C-E Italy (Majella Natl. Park)	AY226837

The classificaton follows Schwarz (1936–1939, 1964). ITS sequence of *Q. morisii* (*Q. suber* × *Q. ilex*) was not determined and is not deposited in the GenBank.

IIntraspecific sampling focused on those taxa with largest distributions. Total genomic DNAs from individual plants were extracted from green tissues (leaves and buds) using the DNeasy Plant minikit (Qiagen), according to the manufacturer's instructions. The 5.8S nrDNA and flanking ITS1 and ITS2 regions were PCR amplified on a GeneAmp2400 (Perkin–Elmer) with 5′-CGTAAC AAGGTTTCCGTAGG-3′ and 5′-AGCGGGTAGTC CCGCCTGA-3′ as primers, determined on the basis of the nucleotide sequence of *Populus deltoides* 18S-25S rRNA region (D'Ovidio, 1992). PCRs were carried out with 10 ng of genomic DNA in 25 μl volume with the puRe Taq Ready-To-Go PCR Beads (Amersham–Pharmacia), following the manufacturer's instructions. The thermal cycling profile consisted in an initial denaturation step of 1 min at 98 °C, followed by 35 cycles of 30 s at 98 °C, 1 min at 58 °C, 1 min at 72 °C and a final extension step of 5 min at 72 °C. Clear cut single-banded fragments were visualised on 1% agarose gels, purified using Nucleospin Extract (Macherey & Nagel) and directly sequenced in both directions by using the amplification primers. Sequence overlaps were confirmed by using 5′-TCGGCAACGGATATCTCGGC-3′ and 5′-ATCCGT TGCCGAGAGTCGT-3′ as forward and reverse primers, respectively, determined upon sequence analysis of oak 5.8S region. Cycle Sequencing and the BigDye Terminator Ready Reaction Kit (Applied Biosystems) were used. Data were collected on ABI Prism 373A automated gel reader. Resulting ITS sequences were further checked by eye with the software CHROMAS 1.43 (C. McCarthy, School of Biomolecular and Biomedical Sciences, Brisbane, Australia), assembled and aligned for several standard descriptive parameters (including size, percentage G + C content, restriction analysis, percentage pairwise divergence) with the Sequence Analysis Software DNAMAN-1999 (Lynnon Biosoft). ITS amplified region of *Q. morisii*, a supposed hybrid between *Q. ilex*, and *Q. suber*, and those from the two parental species underwent a discriminative restriction analysis by use of endonucleases XMN I and DRA III. Digestions were performed according to the protocols supplied by the manufacturer (New England Biolabs) following Nucleospin purification.

2.2. Secondary structures

The boundaries of the internal transcribed spacers (ITS1, ITS2) and 5.8S coding region were determined by comparison with published sequences of angiosperms (among the others: Baldwin et al., 1995; D'Ovidio, 1992; Herschkowitz and Lewis, 1996; Manos et al., 1999).

Predicted secondary structures of the ITS1 and ITS2 RNA transcripts and associated free energy values were evaluated with the minimum-free energy (MFE) algorithm (Zuker, 1989), together with the latest free energy rules (Mathews et al., 1999; Zuker et al., 1999). Fold predictions were made by use of the mfold version 3.1 by Zuker and Turner (www.bioinfo.rpi.edu/applica tions/mfold). Folding temperature was fixed at 37 °C and a search within 5% of the thermodynamic optimality set was performed. The structure annotation based on the heuristic descriptor P-num (propensity of individual bases to participate in base pairs; Zuker and Jacobson, 1998) and constraints to prohibit a string of consecutive bases from pairing (5′ and 3′ ends of transcripts) were used. Finally, predicted structures were

visualised and edited by the RnaViz 2.0 package (De Rijk et al., 2003).

2.3. Phylogenetic analysis

Outgroup taxa (*Colombobalanus excelsa* Nixon & Crepet and *Trigonobalanus verticillata* Forman, GenBank Accession Nos. AF098412 and AF098413) were chosen on the basis of previous phylogenetic studies within *Fagaceae* (Crepet and Nixon, 1989; Manos et al., 1999; Nixon and Crepet, 1989). Other taxa were considered as possible outgroups among *Castanea*, *Castanopsis* and *Lithocarpus* (after Manos et al., 2001). Anyway, ITS bootstrap support for *Quercus* s.l. in the cited work was lower than 50%, hence we preferred to use more external outgroups.

Optimal multiple alignment was obtained with CLUSTALW 1.81 (Thompson et al., 1994) and checked by eye. The secondary structure models were used to resolve ambiguous alignments, as already reported in previous studies (for instance in Gottschling et al., 2001). Parsimony analysis was performed with PAUP* 4.0b1 (Swofford, 1998). A branch-and-bound search was run with default options. All characters were weighted equally and character state transitions were treated as unordered. Gaps were treated as "simple indel coding" after Simmons and Ochoterena (2000), coding them with the software Gapcoder (Young and Healy, 2003). This process codes indels as separate characters, regardless of length, in a data matrix, which is then considered along with the DNA base characters in the phylogenetic analysis.

A maximum likelihood (Felsenstein, 1981) search approach was done as follows: we used Modeltest 3.06 (Posada and Crandall, 1988) to evaluate the likelihood of 56 different models of sequence evolution on the basis of our data. The likelihood ratio test option was used to compare likelihood scores in a nested design and the most likely model of evolution was used as setting in a maximum likelihood (ML) phylogenetic analysis in PAUP* 4.0b1. The maximum likelihood heuristic search was done with 10 random additions and TBR branch swapping. Moreover, the likelihood value of each of the four previously obtained branch-and-bound parsimony trees was calculated. Bootstrap (Felsenstein, 1985) resampling was performed using TBR branch-swapping with 10 random taxon entries per replicate and MULTREES option in effect (with 1000 replicates) under parsimony criterion. MacClade version 3.1 (Maddison and Maddison, 1992) was finally used to trace and map character states onto the consensus tree and to evaluate less parsimonious positions of some taxa. We tested the significance of alternative phylogenetic hypotheses with the Templeton test (Templeton, 1983), as implemented in PAUP* 4.0b1. The congruence of ITS1 and ITS2 data sets was evaluated using the incongruence-length difference (ILD) test of Farris et al. (1995) using PAUP* 4.0b1.

We also examined the phylogenetic position of Italian oaks in a broader context of *Quercus*. A preliminary heuristic search with 10 random addition replicates and TBR branch swapping was done. Gaps were treated as simple indel coding (see above). Bootstrap resampling was performed using faststep option in PAUP* 4.0b1 with 1000 replicates under parsimony criterion. We included in the matrix species of *Quercus* chosen to take into account all relevant infrageneric taxa of the genus. GenBank accessions for these species are available in Manos et al. (1999, 2001).

3. Results

3.1. ITS sequence analysis

The higher denaturing PCR conditions and more stringent annealing temperatures than those usually reported for oak ITS (cf. Muir et al., 2001; Samuel et al., 1998) allowed us to recover single clear-cut bands of ca. 600 bp from all the *Quercus* taxa. Nucleotide sequences of the internal transcribed spacers (ITS1 and ITS2) and the 5.8S coding region of nuclear ribosomal DNA repeats were obtained for 16 individuals, representing 12 of the main Italian *Quercus* species, by means of PCR product direct sequencing. A BLAST (Altschul et al., 1997) search was performed to exclude the sequencing of any contaminant organism. With no exception, the most significant alignments were detected with ITS sequences from *Quercus* co-species (if present in the GenBank) or with other *Quercus* taxa. Upon sequencing, the ITS amplification product of *Q. morisii* displayed coexistence of different repeat units; therefore, analysis of this individual was addressed to PCR-RFLP.

The 15 ITS sequences of oaks presented only small length differences and could be aligned unambiguously, producing a consensus sequence of 600 bp in length. In ITS1, three large deletions were identified: one (7 bp) was common to *Q. crenata* and to both samples of *Q. suber*; a second one (8 bp) was shared by *Q. frainetto*, *Q. robur* and *Q. petraea*. The third (7 bp) was detected only in *Q. trojana*. In ITS2, one insertion of 7 bp was found in sample *Q. ilex*1.

The absolute size of the ITS1+5.8S+ITS2 region (without gaps) ranged from 582 to 600 bp; ITS1 extended between 214 and 222 bp; ITS2 from 205 to 215 bp. The 5.8S coding region showed no size heterogeneity (163 bp). The mean G+C content in the entire ITS region (ITS1+5.8S+ITS2) of the *Quercus* species was 63.9% with a higher value in ITS2 (68.8%) than in ITS1 (66.7%). These values fit within the known ranges for flowering plants, as reported in Baldwin et al. (1995) and are very similar to those detected in *Quercus* spp. by Manos et al. (1999) but not to those obtained by other research groups (Muir et al., 2001; Samuel et al., 1998; data not shown).

Values of pairwise sequence divergence ranged from 0.9 to 5.4% among *Quercus* species and from 7.7 to 10.7% between *Quercus* and the outgroups. Four species, *Q. ilex*, *Q. cerris*, *Q. suber*, and *Q. pubescens*, were represented by two individuals each and showed some intraspecific divergence (0.7, 0.5, 0.2, and 1.2%, respectively). As regards to the species in common with previous studies, we detected up to 0.9% sequence divergence for the entire ITS region in *Q. suber*, 1.2% in *Q. ilex* and in *Q. coccifera*, 1.9% in *Q. cerris*, and 2.6% in *Q. robur* (all sequences from Manos et al., 1999). A higher divergence (>10%) was displayed by alignments with ITS1 + ITS2 sequences from Samuel et al. (1998) and with 5.8S + ITS2 from *Q. robur* and *Q. petraea* functional paralogues (3.4–8.8%), reported by Muir et al. (2001).

3.2. ITS sequences in hybrid species

Quercus crenata is an auto-allogamous, stabilised hybrid (or hybrispecies, *sensu* Nardi, 1988) between *Q. cerris* and *Q. suber* (Bellarosa et al., 1996), and *Q. morisii* is a presumed hybrid between *Q. suber* and *Q. ilex*. The electropherograms of *Q. crenata* clearly showed the presence of a single sequence repeat which shared a higher identity with the ITS sequences of *Q. suber* (99.3%) rather than with those of *Q. cerris* (98.8%).

In contrast, *Q. morisii* was characterised by at least two different repeat types. The eye-check of its sequence electropherograms displayed several double peaks, easily assignable either to *Q. ilex* and to *Q. suber* sequences, i.e., the two presumed parental species. Thus, based on the differential restriction sites of their sequences, a PCR-RFLP analysis was performed on the three genotypes and, upon digestion, the ITS region of *Q. morisii* showed the contemporary presence of the *Q. ilex* and *Q. suber* restriction patterns (Fig. 1).

3.3. ITS1 and ITS2 secondary structure in oaks

Thermodynamically, the secondary structure of ITS1 and ITS2 RNA transcripts showed a fairly uniform stability, as inferred from the lower free energy values, ranging from −89.3 to −107.6 (kcal/mol) and from −96.4 to −102.9 (kcal/mol), for the two spacers, respectively (Table 2). These data are in agreement with the accuracy tests of secondary structures of rRNAs (Mathews et al., 1999), with the only published study involving ITS transcripts analysis of oak species (Mayol and Rossellò, 2001) and with what resulted from a computational analysis that was performed as a check on several *Quercus* ITS sequences deposited in GenBank. As in Mayol and Rossellò (2001), ITS1 from *Q. ilex* and *Q. suber* showed similar, sensibly lower absolute values of free energy than the other oak species.

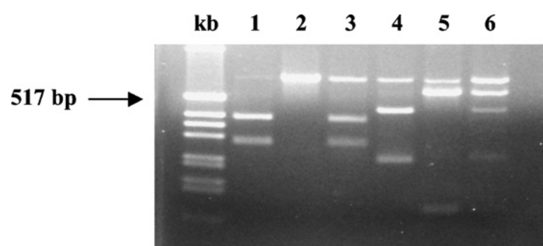

Fig. 1. Restriction analysis of ITS regions amplified from *Q. ilex* (lanes 1 and 4), *Q. suber* (lanes 2 and 5), and *Q. morisii* (lanes 3 and 6) with XMN I (lanes 1–3) and DRA III (lanes 4–6) enzymes. The bands produced in *Q. ilex* and *Q. suber* are simultaneously present in *Q. morisii* patterns. kb = 1 kb DNA ladder.

Foldings between 5% from the minimum free energy with the superimposition of constraints to prohibit 5′ and 3′ ends to pair yielded 1 or 2 folding predictions for every transcript. The criteria for deriving a set of mutually compatible models for oak ITS1 and ITS2 secondary structures were based upon the less free energy values required for the folding, the relative similarity of models between, at least, closely related taxa and the common presence of all the presently known conserved substructures in angiosperms. The first optimal (the second suboptimal in the limited case of *Q. ilex* ITS1) secondary structures always fulfilled these requirements.

Every ITS model showed to possess absolutely uniform fold predictions, which conform well with the known substructural features of algal and vascular plants ITS transcripts (Herschkowitz and Zimmer, 1996; Mai and Coleman, 1997). As shown in Fig. 2, all predicted ITS1 secondary structure presents a common model with 6 helices departing from a single-stranded region. Only small variation in length and loop size (helices III–V) could be detected, that seemingly reflect deeper level relationships within *Quercus*. No extra helices occur, contrary to the sample of *Q. coccifera* examined by Mayol and Rossellò (2001). The Angiosperm Universal Core motif (GGCRY-$(n)_{4-7}$-GYGYC AAGGAA) of Liu and Schardl (1994) was detected in the ITS1 of all the oak species and is invariably present in helix V, in agreement with known predictions: 15 bases are stem paired, whereas the remaining AAGGA A is not. It is perfectly conserved in size (21 bp) and sequence with the exception of *Q. ilex* (both samples) and *Q. coccifera* where the first cytosine is substituted by a thymine. It is noteworthy that the same substitution is present in *Q. ilex* and *Q. coccifera* ITS1 sequenced by Manos et al. (1999) and preserves the stem pairing, thus showing a hemi-compensatory base change behaviour. Helix VI is the longest stem–loop structure and the relative sizes and succession of the other helices is maintained across all taxa. Observation of the folding of helices III and IV served as a guideline for gaps positioning.

The structure of ITS2 models in Fagaceae has been recently described by Coleman (2003) and unambiguously conforms with the common model of all taxa in the present

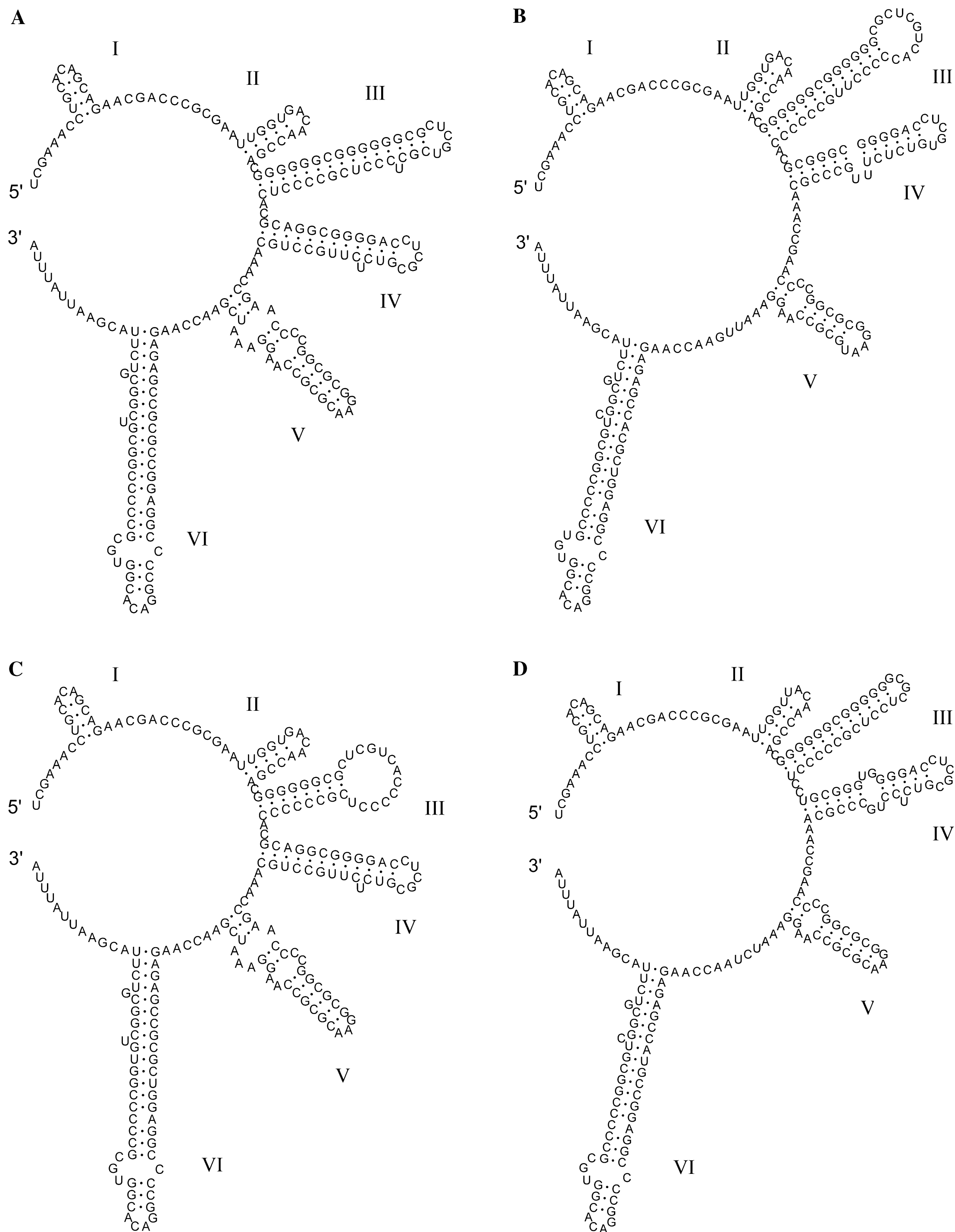

Fig. 2. Putative secondary structures of the ITS1 portion of the primary RNA transcripts in oak species. The predicted conformations present highly uniform foldings that can be represented by four models with slight differences. Oak species sharing the same folding are indicated with parenthesis. (A) *Q. cerris* (*Q. macrolepis*, *Q. trojana*), (B) *Q. coccifera* (*Q. ilex*); (C) *Q. suber* (*Q. crenata*); (D) *Q. petraea* (*Q. pubescens*, *Q. robur*, *Q. frainetto*).

study, with the conserved 4-helix domain that emerged across a wide spectrum of other angiosperms ITS2 from thermodynamic predictions (Mai and Coleman, 1997). Helix II contains the universally conserved pyrimidine bulge; the conserved UGGU sequence is present near the apex of helix III, on the 5′ side. Helix III is also the longest stem–loop substructure and the single stranded sequences connecting the four helices are rich in conserved adenine residues. Fig. 3 shows the high uniformity of models, with only helix III responsible for some length variation in *Q. petraea*-*robur*-*frainetto*-*pubescens*. One extra helix, as in the case of *Q. ilex* examined by Mayol and Rosselló (2001), was found only in one sample (*Q. ilex* 1) and is promoted by the insertion of 7 bp in the spacer sequence.

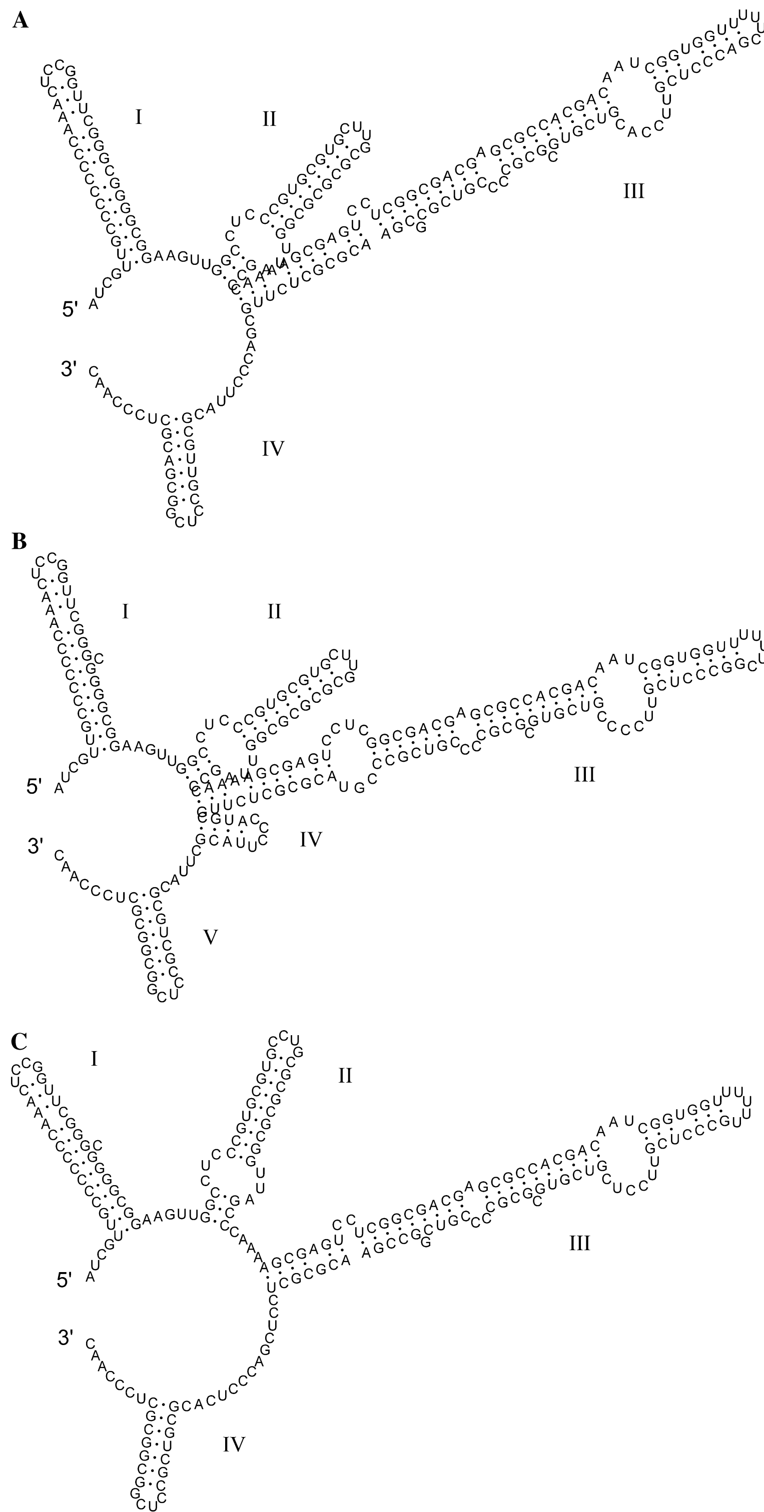

Fig. 3. Putative secondary structure of the ITS2 RNA transcripts. Oak individuals sharing the same folding are indicated with parenthesis. (A) *Q. cerris* (*Q. coccifera*, *Q. ilex* 2, *Q. macrolepis*, *Q. trojana*, *Q. crenata*, *Q. suber*); (B) *Q. ilex* 1, presenting an extra helix; (C) *Q. petraea* (*Q. robur*, *Q. pubescens*, *Q. frainetto*).

Reconcilement of secondary structures with oak ITS2 sequence conservation reflects the model proposed by Herschkowitz and Zimmer (1996) who showed that, among flowering plants, the ITS2 region is characterised by the presence of a complex pattern of conserved (c1–c6) and variable (v1–v6) domains. As a result, gaps positioning in sequences alignment was favoured. Helix I is made up by the pairing between conserved motifs c1 and

Table 2
Analysis of the sequence structure and thermodynamic parameters of ITS1 and ITS2 regions in 11 species (15 individuals) of *Quercus*

Taxa	ITS1			ITS2							
	Paired conserved region (bp)	Six-helix model	ΔG^0 (37 °C) (kcal/mol)	Sequence length of the conserved domains (bp)						ΔG^0 (37 °C) (kcal/mol)	Four-helix model
				C1	C2	C3	C4	C5	C6		
Q. cerris 1	21	Ok	−107.2	14	25	33	20	10	5	−101.0	Ok
Q. cerris 2	21	Ok	−107.2	14	25	33	20	10	5	−101.0	Ok
Q. crenata	21	Ok	−93.6	14	$25^{(T/C)a}$	33	20	10	5	−102.9	Ok
Q. macrolepis	21	Ok	−106.9	14	$25^{(T/C)a}$	33	20	10	5	−102.9	Ok
Q. suber 1	21	Ok	−89.4	14	$25^{(T/C)b,a}$	33	20	10	5	−102.2	Ok
Q. suber 2	21	Ok	−89.4	14	25	33	20	10	5	−101.0	Ok
Q. trojana	21	Ok	−107.6	$14^{(T/C)}$	$25^{(G/A;T/C)a}$	33	20	10	5	−100.7	Ok
Q. frainetto	21	Ok	−102.6	14	25	33	20	$10^{(T/C)a}$	5	−100.3	Ok
Q. petraea	21	Ok	−102.4	14	25	33	20	10	5	−98.3	Ok
Q. pubescens 1	21	Ok	−106.5	14	25	33	20	10	5	−98.5	Ok
Q. pubescens 2	21	Ok	−99.1	14	25	33	20	$10^{(T/C)}$	5	−98.5	Ok
Q. robur	21	Ok	−99.1	14	25	33	20	$10^{(T/C)a,b}$	5	−96.4	Ok
Q. coccifera	$21^{(T/C)a,b}$	Ok	−99.7	14	25	33	20	10	5	−100.8	Ok
Q. ilex 1	$21^{(T/C)a,b}$	Ok	−89.3	14	25	33	20	10	$5^{(C/G)}$	−102.5	Ok (extra helix)[a]
Q. ilex 2	$21^{(T/C)a,b}$	Ok	−89.3	14	25	33	$20^{(G/C;G/A)b}$	10	5	−99.8	Ok

Nucleotide substitutions in putative conserved domains (c1–c6) and the relative correspondence of the predicted secondary structures to a common model are indicated.

[a] Same substitutions.

[b] Mutations shared with the species in common with Manos et al. (1999).

part of c2; remaining c2 and c3 pair to constitute Helix II; c4 and c5 contribute to helix III, whereas helix IV is made up by hypervariable segment v6. The six conserved motifs showed very high uniformity in size, sequence and position in all taxa. A few exceptions were found and they are presented in Table 2. Interestingly, most of such substitutions are conserved among closer species and keep the pairing state (T/C and G/A substitutions with G and U, respectively).

3.4. Phylogenetic analysis

Together with the outgroups, alignment of the final matrix required the introduction of 16 indels, of which 10 were 1-bp long, distributed throughout ITS1 and ITS2. The parsimony data matrix contained 475 constant, 40 parsimony-uninformative, and 93 parsimony-informative characters. The branch-and-bound search produced four maximum parsimony trees of 153 steps, with consistency index (CI) = 0.824 and retention index (RI) = 0.852.

The simplest maximum likelihood model identified with Modeltest 3.06 for our data assumed equal base frequencies, six substitution categories and gamma distributed rate heterogeneity, partitioned into four rate categories. These settings correspond to the General Time Reversible Model, GTR+G (Huelsenbeck and Crandall, 1997; Posada and Crandall, 1988).

One of the four maximum parsimony trees, corresponding to that with the best likelihood value after the chosen substitution model is visualised in Fig. 4. The bootstrap values (1000 replicates) are reported. The Templeton test was used to evaluate the significance of alternative phylogenetic hypotheses arising from the MP trees, or less parsimonious solutions suggested by the known literature.

The maximum likelihood analysis (that cannot include information provided by indels) produced one tree which was largely congruent with the topology in Fig. 4, but was less resolved with two polytomies: one at the base of the "*Cerris*" group and the other at the base of the ingroup, with *Q. ilex* + *Q. coccifera* unresolved with respect to the other two clades (data not shown).

The monophyly of the ingroup was corroborated by 100% bootstrap. Two main lineages emerged from the root, with high bootstrap support: a clade corresponding to *Q. petraea*, *Q. frainetto*, *Q. robur*, and *Q. pubescens*, here indicated as "*Quercus* group" and its sister group formed by *Q. coccifera* and *Q. ilex*, corresponding to "*Ilex* group," plus *Q. cerris*, *Q. crenata*, *Q. suber*, *Q. macrolepis*, and *Q. trojana*, corresponding to the "*Cerris* s.s. group."

In the "*Quercus* group," *Q. pubescens* were sister to the group formed by *Q. petraea*, *Q. frainetto* and *Q. robur* (73% bootstrap support), while *Q. frainetto* was the closest relative to *Q. robur* (66% support). Testing alternative phylogenies with MacClade indicated that forcing *Q. pubescens* with *Q. petraea* produced a tree of 154 steps (only one step longer than the most parsimonious one), with a CI = 0.82 and an RI = 0.85 and Templeton test indicating a not significant difference as compared with the most parsimonious solution; grouping *Q. pubescens* with *Q. frainetto* implied 2 further steps

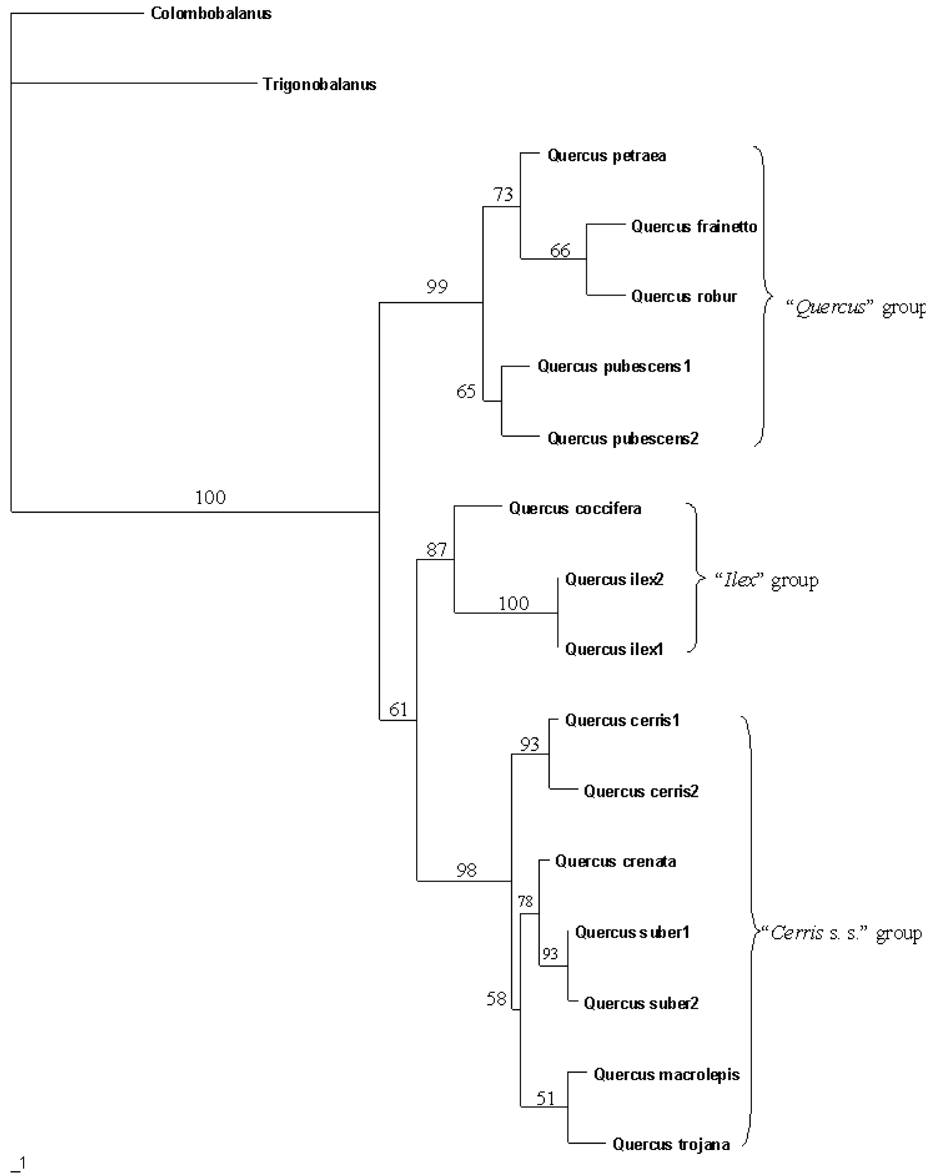

Fig. 4. One of the four maximum parsimony trees based on nrDNA ITS sequences, corresponding to that with the best maximum likelihood value. Gaps were treated as "simple indel coding" after Simmons and Ochoterena (2000). Bootstrap values (1000 replicates) are reported.

(CI = 0.81, RI = 0.83, not significant difference) while *Q. pubescens* with *Q. robur* originated a maximum parsimony tree of 158 steps (CI = 0.80, RI = 0.83, significant difference with the Templeton test).

The "*Ilex* group" and the "*Cerris* s.s. group" resulted as sister groups in our phylogenetic reconstruction. In the "*Cerris* s.s. group," corroborated by 98% bootstrap support, *Q. cerris* was in basal position. The closest relative of *Q. suber* was *Q. crenata*, while *Q. macrolepis* and *Q. trojana* (unresolved in two MP trees, data not shown) clustered together with a low bootstrapping value, and resulted as sisters to the former two species. With MacClade it was possible to see that forcing *Q. cerris* with *Q. crenata*, *Q. macrolepis* and *Q. trojana* produced an only one-step longer maximum parsimony tree in all the three cases (154 steps, CI = 0.82, and RI = 0.85, not significant difference with Templeton test). *Q. crenata* in an intermediate position between *Q. suber* and *Q. cerris* produced a maximum parsimony tree (153 steps, CI = 0.82, RI = 0.85). Grouping *Q. cerris* with *Q. trojana* and *Q. suber* with *Q. macrolepis* implied only one step more than the most parsimonious hypothesis but with a significant difference after the Templeton test. A 10-steps longer tree (highly significant difference) was obtained when *Q. suber* was forced with *Q. coccifera* and *Q. ilex* (forming an "evergreen" clade).

Analysing ITS1 and ITS2 separately indicated that in *Quercus* species ITS1 sequence is more informative than ITS2. In fact, 27 characters were informative in ITS1, 4 in 5.8S, 25 in ITS2 excluding gaps. The phylogenetic

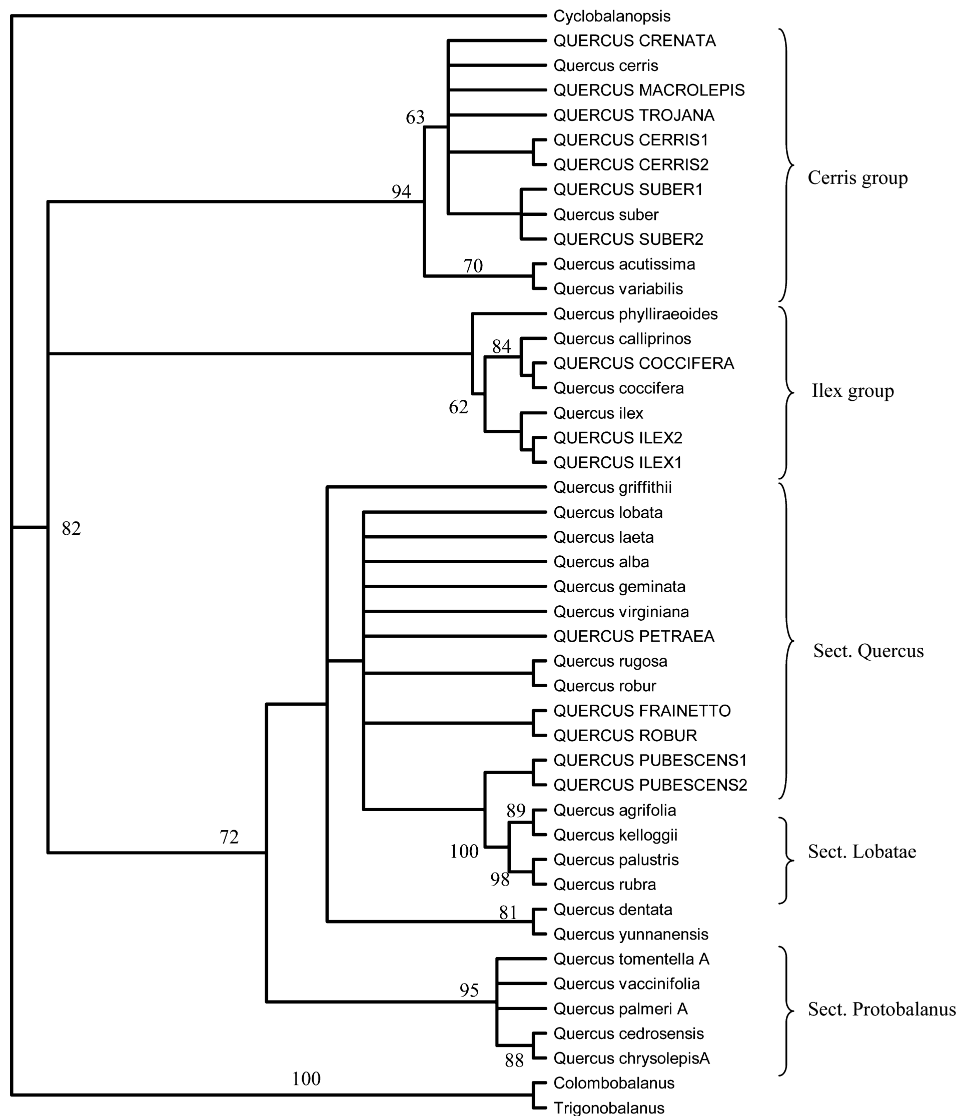

Fig. 5. Consensus tree of 3064 maximum parsimony trees produced with an analysis of ITS sequences of these study (indicated in capital letters) together with species representative of all infrageneric groups in *Quercus*. Bootstrap values are reported on branches if higher than 50%.

information obtained from ITS1 was sufficient to produce the phylogenetic reconstruction obtained with the whole ITS sequence. The ITS2 alone didn't show any conflict with the ITS1 phylogeny, but it could not adequately resolve the ingroup (data not shown). However, the ITS1 and ITS2 data sets (excluding indels) were found to be combinable according to the ILD test ($P = 0.773$).

The preliminary analysis of Italian oaks in a broader context of *Quercus* done with the ITS sequences newly derived by the authors and most species of the genus available on GenBank produced a matrix with 613 characters (excluding gaps), 3064 maximum parsimony trees with $L = 396$ steps, $CI = 0.631$ and $RI = 0.798$ (consensus tree in Fig. 5). *Quercus robur-petraea-frainetto-pubescens* clustered together within the *Quercus* s.s. group described by Manos et al. (1999); all other Italian oaks were in two monophyletic clades, the "*Ilex*" and "*Cerris*" groups, that closely resemble subg. *Cerris* and subg. *Schlerophyllodrys*, as described by Schwarz (1936–1939, 1964).

4. Discussion

4.1. Utility of ITS data in oak phylogeny

The present study well supports the ITS region as a useful nuclear marker in phylogenetic reconstructions of Italian oaks. Our paper is set in a context (woody genera) where there is still limited knowledge for routine application of new markers (e.g. single-copy nuclear genes). Conversely, organellar DNA variation appeared to be less informative, giving rise, in some

instances, to phylogenetic incongruence (see Manos et al., 1999).

Previous studies on oak phylogeny using ITS highlighted the recurrent possibility of producing distorted reconstructions, due to unnoticed intra-individual polymorphism and the consequent contemporary analysis of divergent paralogues. Apart from large ITS cloning approaches, a series of alternative strategies can be used to test hypotheses of sequence paralogy or orthology, like evaluation of patterns of nucleotide substitution (e.g. relative divergence of regions and subsequences, length variation, G + C content), estimates of stability, and uniformity of secondary structures of transcripts (Bailey et al., 2003). Based on such analyses, our study generated highly uniform sequences, with high probability of belonging to a single class of ITS paralogues and for which it is conceivable to assume functionality. Furthermore, they could be easily incorporated in oak molecular data sets that largely agree with the most recent taxonomic schemes (e.g., that in Manos et al., 2001).

Our results allowed a clear distinction of all the oak species, even when more than one individual per species were investigated. Indeed, analyses of the entire ITS region (ITS1+5.8S + ITS2) and the avoidance of interferring paralogues were determinant. Such good resolution may be due also to the geographical origin of the material we tested, and to the well-aimed choice of the collecting sites. In fact, all the species were sampled in Italy, which, as previously pointed out (Follieri et al., 1986; Spada et al., 1996) and recently confirmed at the molecular level (Petit et al., 2002a,b), constitutes one of the three Mediterranean peninsulas that represented the major refugial areas for the European vegetation during the last Ice Age. In particular, two sampling areas (SE Italy and Latium) are to be seen as Würm refugia, based on previous paleobotanical and vegetational studies (cf. Brewer et al., 2002; Follieri et al., 1998; Schirone and Spada, 1995; Spada et al., 1996). These are variously differentiated territories, in an ecological sense, located at mid altitudes in mountainous areas, that provided temporary refugia from climatically adverse conditions, and allowed the survival of small and isolated populations of trees. Here the species could safeguard their own identities before spreading north- and westwards, at the end of glaciations. Oak populations sampled in such areas, when investigated, have already demonstrated to constitute well differentiated genepools with rich genetic diversity (for instance in Fineschi et al., 2002), whereas pollen flow and local human management have been invoked as evolutionary forces that erased such differentiation in C and N Europe, during recolonisation and merging of the genepools (Petit et al., 2002a).

Besides the validation of ITS sequences as orthologues and the identification of higher order sequences to improve the accuracy of alignment, information from secondary structures predictions were used to explore possible phylogenetic implications. The difficulty in tracking organismal relationships from the analysis of putative ITS secondary structure is emphasised by a level of sequence conservation that is too great to be structurally revealing within a group of close relatives as well as by the lack of a more in-depth study of intraspecific oak ITS variability. Nevertheless, the recurrent, though minimal, substructure differences among groups were congruent and often complementary with traditional *Quercus* taxonomy and molecular information from ITS data sets.

4.2. Section *Quercus*

In our phylogenetic reconstruction, one of the two main lineages depicting the ITS profile of Italian oaks is that constituted by *Q. petraea*, *Q. robur*, *Q. pubescens*, and *Q. frainetto*, all belonging to subg. *Quercus*, sect. *Quercus* s.s. Nixon (subg. *Euquercus*, sections *Lepidobalanus* and *Mesobalanus* Camus; subg. *Quercus*, sections *Robur*, *Roburoides*, and *Dascia* Schwarz), indicated as "*Quercus*" group in Fig. 4. The ITS sequences of these species form a well-differentiated clade, as already anticipated by the use of storage proteins (Bellarosa et al., 1996) and later confirmed by cytogenetic studies (Zoldos et al., 2001). The common deletion in ITS1 (8 bp) of *Quercus frainetto*-*robur*-*petraea* supports the monophyly of this group together with 73% bootstrap values. Moreover, on the basis of both ITS1 and ITS2 secondary configurations, a clear distinction of *Q. petraea*-*robur*-*frainetto*-*pubescens* from the other species is highlighted.

Within this group, *Q. robur* and *Q. petraea* are separated, contradicting the difficulties encountered by Muir et al. (2001) in distinguishing these two species at the molecular level. It is possible that a complete ITS analysis would have highlighted the existing differences, as we showed that most of the informative variation is in ITS1, which was not considered in the above cited study. It is also possible that, as mentioned, sampling in areas other than those considered by Muir and co-workers may have played an important role. Interfertility between these two sympatric species is effectively large and well documented (cf. Petit et al., 2002a; Whittemore and Schall, 1991). The two species have main range in Central Europe, where a confluence of migrating routes from glacial refugia with large evidence of introgression has already been observed. In contrast, in S Europe *Q. robur*-*petraea* are represented by a large number of local differentiations, in which different patterns of genetic flow, introgression and persistence of ancient characters are to be expected. The affinity between *Q. robur* and *Q. frainetto* finds some correspondence in the similar ecological requirements of the two species, that often grow in the same conditions. On the other hand, it contradicts the current classifications (Camus, 1936–1954; Schwarz, 1936–1939) and the data from our previous

investigations carried out on the basis of seed storage proteins (Bellarosa et al., 1996). According to those results, *Q. frainetto* showed a closer relationship with *Q. pubescens*, in agreement with Schwarz's taxonomical scheme where these two species belong to sect. *Dascia* while *Q. robur* belongs to sect. *Robur*. In support of the present data, where grouping *Q. frainetto* with *Q. pubescens* implied two further steps in respect of the most parsimonious reconstruction (with insufficient significance after the Templeton test), some remarkable speculations can be made, on the basis of recent studies by Petit et al. (2002a,b), Bordács et al. (2002), and Fineschi et al. (2002), though with the caution required in comparing data sets which used different molecular markers and, specifically, cpDNA analyses (cf. Manos et al., 1999). The analysis conducted by Petit et al. (2002a), on the haplotypes of European white oaks (*Q. robur*, *Q. petraea*, *Q. frainetto*, and *Q. pubescens*) covering the entire continent, showed that *Q. frainetto* possesses higher affinity (both in type and number of haplotypes) with *Q. robur*. These data are even more outstanding if referred to the same species from the Balkans and Carpathian regions (Bordács et al., 2002). Instead, Fineschi et al. (2002) investigated cpDNA variation in Italian white oaks and found a higher frequency of the same haplotypes in *Q. frainetto* and *Q. pubescens*. It must be noted that our sample of *Q. frainetto* was not collected in C Italy as in the above cited studies (Bellarosa et al., 1996; Fineschi et al., 2002), but in a small relic population from the far SE of Italy (Medagli et al., 1990). Thus, a special affinity between *Q. frainetto* from this region and the Balkans may be assumed, founded on the vegetational southern trans-Adriatic migrations suggested by Schirone and Spada (1995) and reported also by Fineschi et al. (2002). Further investigations on these species system from new geographical areas are obviously required, to uncover the extent of such biogeographical variation. On the basis of ITS data *Q. pubescens* appears to be more distantly related to the other species of the "*Quercus*" group. It must be noted that *Q. pubescens* has a very large distribution area at Mid European latitudes, extending from the Iberian peninsula eastwards to Krimea (Schwarz, 1964); it is the most common white oak in Italy and it is extremely polymorphic, including many different forms whose taxonomic status is particularly difficult to assess (Schirone and Spada, 2001). *Q. pubescens* is extremely frequent in the structure of evergreen or mixed meso-termophilic forests, and it is difficult to say whether this is the heritage of a former aboriginal mixed forest or rather the result of fragmentation due to human impact, that tends to favour the more xerophytic species. In fact, *Q. pubescens* is the most xeromorphic species among Italian *Quercus* s.s. oaks. It probably differentiated from a mesic ancestor within the Arctotertiary oaks, in areas where the long lasting conditions of continental climate favoured the evolution of xeromorphism and the colonisation of S Europe during the glacial peak of Quaternary (Schirone and Spada, 2001).

4.3. Cerris and Ilex groups

The second main lineage of Italian oaks ("*Ilex*" plus "*Cerris*" s.s. group) appears as evolutionarily more ancient, which is consistent with European macrofossil data (Kvacek and Walther, 1989) with the evergreen oaks as the first to appear throughout the continent. This plesiomorphic trait (the evergreen habit) is maintained in *Q. suber*, *Q. ilex*, and *Q. coccifera*, contributing towards justifying their placing in a unique taxonomic group as proposed by some authors (Camus, 1936–1954; Fiori, 1923–1925). However, taxa in this clade are well resolved and our results confirm the findings of Manos et al. (1999), Zoldos et al. (2001), and our previous studies (Bellarosa et al., 1996). This set of oak species constitutes a monophyletic lineage within sect. *Quercus*, corresponding to the *Cerris* group and the *Ilex* group *sensu* Nixon (1993) (subg. *Cerris* and subg. *Schlerophyllodrys* Schwarz). Our data confirm *Q. ilex* and *Q. coccifera* as belonging to the *Ilex* group, and the remaining 5 species (*Q. cerris*, *Q. suber*, *Q. crenata*, *Q. macrolepis*, and *Q. trojana*) to the *Cerris* group. Based on the ITS2 secondary structure, they also constitute a unique group (section *Cerris*). However, some intra-specific variation was reported, as in the case of *Q. ilex* 1 (and *Q. coccifera* in the GenBank), so that the further existence of species-specific subclasses of models which may allow better resolved groupings is possible. With regard to ITS1, on the other hand, the *Ilex* group appears to constitute a differentiated set. The remaining species show highly uniform configurations, with minor difference (caused by a 7 bp common deletion) characterising *Q. crenata* and *Q. suber*, thus contributing to justify the separation between subg. *Schlerophyllodrys* and subg. *Cerris* Schwarz and, within this latter group, between *Q. cerris-trojana-macrolepis* and the evergreen *Q. suber* in different subsections. Furthermore, the hemi-compensatory base changes in the motif of Liu and Schardl (1994) lean towards the proposed separation of the two subgenera.

However, a slight discrepancy with one of our previous results emerged in relation to the *Q. trojana-macrolepis* subclade. Based on IGS restriction patterns (Bellarosa et al., 1990), a closer relationship between *Q. cerris* and *Q. trojana* and *Q. suber* and *Q. macrolepis* was suggested. Forcing these two clades with MacClade produced trees only one step longer than the most parsimonious one (only 51% bootstrap support for this subclade). The data presented here accurately reflect what resulted from the analysis of the storage proteins (Bellarosa et al., 1996) and perfectly match the Camus classification (1936-54), with *Q. trojana* and *Q. macrolepis* in sect. *Cerris*, subsect. *Macrolepides*, whilst *Q. cerris* was placed in the monotypic subsect. *Eucerris* and *Q.*

suber in the subsect. *Suber* of sect. *Cerris*. Alternatively, Schwarz included the four species in different subsections. Indeed, there are strict relationships between *Q. trojana* and *Q. macrolepis* and the two species, from the ecological and distributional viewpoints, share higher similarities than with *Q. cerris* and *Q. suber* (cf. Francini-Corti, 1960). It is worthwhile pointing out that the two species were sampled in the only two unique Italian existing stands, the westernmost of the European range, and that they belong to the tree flora which is likely to have survived in the scattered refugia of Apulia (S-E Italy), where the Salento Peninsula is located. Such isolation of the stands may have constituted a barrier to gene flow, favouring the persistence of more ancient relationships between the species.

It is noteworthy that *Q. cerris* is sister to the clade of the rest of the "*Cerris*" s.s. group, upholding the hypothesis of a common differentiation center of this whole group, probably in the Middle Eastern-PeriCaucasian area (cf. Palamarev, 1989; Schirone and Spada, 2001). Most of the subgenus *Cerris* is semideciduous except for *Q. cerris*, which is fully deciduous. This trait, together with its biennial reproductive cycle, is well suited to overcome stress due to winter cold and allowed the species to migrate northwards (reaching the slopes of the Alps) where it can be found at present. The same route was followed, to a lesser degree, by *Q. trojana* and *Q. macrolepis*, which exploited the same fruit maturation strategy but colonised the southern areas of the Mediterranean basin, with semiarid climate. In this view, deciduousness can be interpreted as a derived character in the subgenus *Cerris*, providing an explanation for the semideciduous habits of most of its taxa.

Quercus suber is a species which substantially overlaps with the *Q. ilex* geographical distribution but it is missing from E Europe. According to Sauvage (1961), *Q. suber* may have originated in the Iberian peninsula, while data from other authors (cfr. Palamarev, 1989 and references therein) tend to suggest a Middle Eastern origin, in common with the whole *Cerris* group, and a S-Western migration in the Mediterranean region, where the character evergreen habit has survived thanks to the lack of climatic constraints. It is noteworthy that, at the geographical extremes of the *Q. suber* distribution, where climatic constraints are present (i.e. winter cold in French Landes or aridity in Apulia), biennial fruit maturation and partial deciduousness can be observed.

Interestingly, when looking at the free energy data of ITS RNA transcripts, *Q. ilex* and *Q. suber* (our samples and those from the GenBank) appear to stand clearly separated from all the other *Quercus* taxa. It is debatable whether this may reflect some physiological peculiarities shared by the two species. In effect, *Q. ilex* and *Q. suber* are the two only evergreen species with similar "oceanic" (humidity demanding) behaviour among the Mediterranean oaks. As mentioned above, the two species often grow in the same environment, form mixed stands and their hybrids (*Q. morisii*) are becoming well documented.

4.4. Hybrid species

Quercus crenata encompasses several introgressive forms between *Q. suber* and *Q. cerris*, to the point that only a statistical analysis of morphological variation may be of help in characterizing the individuals (Bellarosa et al., 1996; Schirone et al., 1990). The most parsimonious analysis determines the position of *Q. crenata* as closely related to the two known parental species and clustering together with *Q. suber* with 78% bootstrap support. We may assume that the hybrid status of *Q. crenata* is accomplishing the rank of a distinct taxon (or hybrispecies), as testified also by its distribution range, which extends to regions where one of the presumed parents is no longer present (e.g. *Q. suber* in the Balkans). It is unclear, however, how completely ITS variation was replaced and whether the two parental ITS loci underwent (i) a process of concerted evolution, being converted to a single sequence variant of ITS, somewhat more similar to that of *Q. suber*, or (ii) interlocus homogenisation, with one parental type gone lost through segregation or still on the way of elimination from the hybrid genome. In effect, Rauscher et al. (2002) demonstrated that direct sequencing may not be able to detect paralogues present in the genome with less than 10% composition, therefore residual sequences related to those of the ancestral parents would be detected only through an extensive cloning approach.

Quercus morisii, instead, clearly exhibited both parental type variants, providing further elements to the ongoing investigations by Toumi and Lumaret (2001), Belahbib et al. (2001), and Lumaret et al. (2002) on the study of introgressive forms between *Q. ilex* and *Q. suber*. It must be noted that this relatively rare hybrid can be found only in presence of both parents. The material we studied was sampled from a small (6 ha) forest on the Salento peninsula in SE Italy, a relic area among vines and olives, with 93% *Q. suber*, 6% *Q. ilex*, and 1% other species. Hybrid individuals were easily recognised due to the morphological traits which were perfectly intermediate between the two parental oaks. The bark, for instance, was neither suberised nor smooth as in *Q. ilex* (cf. Belahbib et al., 2001; Nativitade, 1936). Interestingly, *Q. morisii* plants are quite abundant (0.3%), appearing evenly-aged (ca. 100 years old) with no obvious regeneration, and mostly occur in two well circumscribed areas. The above cited authors stated that there are "probably not significant interspecific genetic exchanges between the two species." We thus assume that these individuals are all F1 hybrids, representing small cohortes originated after peculiar bio-climatic events in this extreme part of the distribution range of *Q. suber*.

5. Conclusion

In this study, we have demonstrated the utility of ITS sequences as a molecular tool to help clarify *Quercus* systematics, which has been highly controversial. In our attempts to examine the taxonomic status of the main groups of Italian oaks, the ITS data we obtained corroborated the molecular phylogeny drawn by Manos et al. (1999) and the taxonomic scheme proposed by Schwarz (1964), which appears to be the most suitable in describing the systematics of European oaks. The topology of the obtained phylogenetic trees highlights the phylogenetic placement of several oak species, including *Q. frainetto*, *Q. pubescens*, *Q. suber*, *Q. trojana*, and *Q. macrolepis.* In our opinion, this set of oak species may deserve further study at the population level for a better understanding of their diversification history, in relation to geographical distribution and shared interspecific relationships.

Finally, our data could not be compared with those from previous studies by Samuel et al. (1998) and Muir et al. (2001), which presented incomplete sequences and/or highly divergent paralogues. Only data sets from Manos et al. (1999, 2001) were consistent with the presented ITS oak sequences and could be efficiently included in our study. Preliminary reconstruction of oak phylogeny based on Italian, N American and Asian species produced a phylogenetic tree in congruence with the geographical origin of the species and the known taxonomic schemes of *Quercus*. The resolution of three main clades supports the recognition of subg. *Schlerophyllodrys*, *Cerris*, and *Quercus*, as proposed by Schwarz (1936–1939, 1964). The geographical split between North American and Eurasian species was generally maintained, apart from the known exception represented by sect. *Quercus*, for which connections between North America and eastern Asia were already suggested (Manos et al., 2001). All the Italian oaks were grouped together with members in the sections where they have been traditionally placed based on morphology, and directly linked to the Eurasian species. This phylogenetic pattern supports the theory of Axelrod (1983) that all major oak lineages evolved locally, in the geographical areas where they differentiated. Axelrod's theory needs to be further tested with additional sampling around the Mediterranean sea and to central Asia.

Acknowledgments

We are grateful to Prof. Fabio Maggini for the critical reading of the manuscript and the fruitful comments, along with his enjoyable friendship and to the two anonymous reviewers whose pieces of advice permitted to us to improve our paper.

Funding for this research was provided by A.R.S.I.A.L. (Regional Office for Agriculture Development, Latium).

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Axelrod, D.I., 1983. Biogeography of oaks in the Arcto-Tertiary province. Ann. Missouri Bot. Gard. 62, 629–657.

Bailey, C.D., Carr, T.G., Harris, S.A., Hughes, C.E., 2003. Characterization of angiosperm nrDNA polymorphism, paralogy and pseudogenes. Mol. Phyl. Evol. 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 on Angiosperm phylogeny. Ann. MO. Bot. Gard. 82, 247–277.

Belahbib, N., Pemonge, M.-H., Ouassu, A., Sbay, H., Kremer, A., Petit, R.J., 2001. Frequent cytoplasmic exchange between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. Mol. Ecol. 10, 2003–2012.

Bellarosa, R., Delre, V., Schirone, B., Maggini, F., 1990. Ribosomal RNA genes in *Quercus* spp. (*Fagaceae*). Pl. Syst. Evol. 172, 127–139.

Bellarosa, R., Schirone, B., Maggini, F., Fineschi, S., 1996. Inter- and intraspecific variation in three mediterranean oaks (*Q. cerris*, *Q. suber*, *Q. crenata*). In: Proceedings of the Workshop on Inter- and Intraspecific Variation in European Oaks: Evolutionary Implications and Practical Consequences, 15–16 June 1994, Bruxelles. Off. Publ. Of the European Communities, Luxembourg, pp. 239–276.

Bennet, K.D., Tzedakis, P.C., Willesk, K.J., 1991. Quaternary refugia of the North European trees. J. Biogeogr. 18, 103–115.

Bordács, S., Popescu, F., Slade, D., Csaikl, U.M., Lesur, I., Borovics, A., Kézdy, P., König, A.O., Gömöröy, D., Brewer, S., Burg, K., Petit, R.J., 2002. Chloroplast DNA variation of white oaks in northern Balkans and in the Carpathian Basin. For. Ecol. Manage. 156, 197–209.

Brewer, S., Cheddadi, R., De Beaulieu, J.-L., Relle, M., Data contributors, 2002. The spread of deciduous *Quercus* throughout Europe since the last glacial period. For. Ecol. Manage. 156, 27–48.

Burger, W.C., 1975. The species concept in *Quercus*. Taxon 24, 45–50.

Camus, A., 1936-1954. Les Chenes. Monographie du genre *Quercus* and monographie du genre *Lithocarpus*. Encycl. Econ. de Sylvic., Vol. VI, VII, VIII. Lechevalier, Paris.

Coleman, A.W., 2003. ITS2 is a double-edged tool for eukaryote evolutionary comparisons. Trends Genet. 19, 370–375.

Crepet, W.L., Nixon, K.K., 1989. Earliest megafossil evidence of *Fagaceae*: phylogenetic and biogeographic implications. Am. J. Bot. 76, 842–855.

D'Ovidio, R., 1992. Nucleotide sequence of a 5.8S rDNA gene and of the internal transcribed spacers from *Populus deltoides*. Plant Mol. Biol. 19, 1069–1072.

De Rijk, P., Wuyts, J., De Wachter, R., 2003. RnaViz2: an improved representation of RNA secondary structure. Bioinformatics 19, 299–300.

Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995. Testing significance of incongruence. Cladistics 10, 315–319.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Ferris, C., Oliver, R.P., Davy, A.J., Hewitt, G.M., 1993. Native oak chloroplast reveal an ancient divide across Europe. Mol. Ecol. 2, 337–344.

Fineschi, S., Taurchini, D., Grossoni, P., Petit, R.J., Vendramin, G.G., 2002. Chloroplast DNA variation of white oaks in Italy. For. Ecol. Manage. 156, 103–114.

Fiori, A., 1923–1925. Nuova Flora analitica d'Italia. Edagricole, Bologna, Italy.

Follieri, M., Magri, D., Sadori, L., 1986. Late Pleistocene *Zelkova* extinction in Central Italy. New Phytol. 103, 269–273.

Follieri, M., Giardini, M., Magri, D., Sadori, L., 1998. Palynostratigraphy of the last glacial period in the volcanic region of Central Italy. Quatern. Int. 47–48, 3–20.

Francini-Corti, E., 1960. Aspetti evolutivi del ciclo ontogenetico nella sistematica dei generi *Pinus* e *Quercus*. Relazioni e discussione del colloquio internazionale. In: Evoluzione e genetica. Quad. Acc. Naz. Lincei, vol. 60, pp. 71–103.

Gottschling, M., Hilger, H.H., Wolf, M., Diane, N., 2001. Secondary structure of the ITS1 transcript and its application in a reconstruction of the phylogeny of Boraginales. Plant Biol. 3, 629–636.

Herschkowitz, M.A., Lewis, L.A., 1996. Deep-level diagnostic value of the rDNA-ITS region. Mol. Biol. Evol. 13, 1276–1295.

Herschkowitz, M.A., Zimmer, E.A., 1996. Conservation patterns in Angiosperm rDNA-ITS2 sequences. Nucleic Acids Res. 24, 2857–2867.

Hewitt, G.M., 1996. Some genetic consequences of ice ages and their role in divergence and speciation. Biol. J. Linn. Soc. 58, 247–276.

Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Systematics 28, 437–466.

Kubitzki, K., 1993. Fagaceae. In: Kubitzki, K., Rohwer, J.G., Bittrich, V. (Eds.), The families and genera of vascular plants, vol. 2. Springer, Berlin, pp. 301–309.

Kvacek, Z., Walther, H., 1989. Paleobotanical studies in *Fagaceae* of the European Tertiary. Plant Syst. Evol. 162, 213–299.

Liu, J.-S., Schardl, C.L., 1994. A conserved sequence in internal transcribed spacer 1 of plant nuclear rRNA genes. Plant Mol. Biol. 26, 775–778.

Lumaret, R., Mir, C., Michaud, H., Raynal, V., 2002. Phylogeographical variation of Chloroplast DNA in holm oak (*Quercus ilex* L.). Mol. Ecol. 11, 2327–2336.

Maddison, W.P., Maddison, D.R., 1992. MacClade: Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, USA.

Mai, J.C., Coleman, A.W., 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green *Algae* and flowering plants. J. Mol. Evol. 44, 258–271.

Manos, P.S., Doyle, J.J., Nixon, K.C., 1999. Phylogeny, biogeography and processes of molecular differentiation in *Quercus* subgenus *Quercus* (*Fagaceae*). Mol. Phyl. Evol. 12, 333–349.

Manos, P.S., Zhou, Z.-K., Cannon, C.H., 2001. Systematics of *Fagaceae*: phylogenetic tests of reproductive trait evolution. Int. J. Plant Sci. 162 (6), 1361–1379.

Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.

Mayol, M., Rossellò, J.A., 2001. Why nuclear ribosomal DNA spacers (ITS) tell different stories in *Quercus*. Mol. Phyl. Evol. 19, 167–176.

Medagli, P., Bianco, P., Schirone, B., D'Emerico, S., Ruggiero, L., 1990. Il farnetto di Bosco Belvedere (Lecce). Ann. Bot. (Roma) XL VIII (Suppl. 7), 77-83.

Muir, G., Fleming, C.C., Schlötterer, C., 2001. Three divergent rDNA clusters predate the species divergence in *Quercus petraea* (Matt.) Liebl. and *Quercus robur* L. Mol. Biol. Evol. 18, 112–119.

Nardi, E., 1988. Reflections on the genetic bases of the concept of "species". Inf. Bot. Ital. 20, 159–192 (in Italian).

Nativitade, J.V., 1936. Estudo histològico dos peridermes do hibrido *Quercus ilex* × *Quercus suber*. P. Cont. Pub. Direcção General Florestas et Aquicolas te: 343–468.

Nixon, K.C., 1993. Infrageneric classification of *Quercus* (*Fagaceae*) and typification of sectional names. Ann. Sci. Forest. 50 (Suppl. 1), 25s–34.

Nixon, K.C., Crepet, W.L., 1989. *Trigonobalanus* (Fagaceae): taxonomic status and phylogenetic relationships. Am. J. Bot. 6, 828–841.

Palamarev, E., 1989. Paleobotanical evidences of the Tertiary history and origin of the Mediterranean sclerophyll dendroflora. Pl. Sist. Evol. 162, 93–107.

Petit, R.J., Csaikl, U.M., Bordács, S., Burg, K., Coart, E., Cottrell, J., van Dam, B.C., Deans, J.D., Fineschi, S., Dumolin-Lepègue, S., Finkeldey, R., Gillies, A., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Mátiás, G., Munro, R.C., Pomonge, M.-H., Popescu, F., Slade, D., Tabbener, H., Taurchini, D., de Vries, S.M.G., Ziegenhagen, B., Kremer, A., 2002a. Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. For. Ecol. Manage. 156, 5–26.

Petit, R.J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U.M., van Dam, B.C., Deans, J.D., Fineschi, S., Finkeldey, R., Glaz, I., Goicoechea, P.G., Jensen, J.S., König, A.O., Lowe, A.J., Madsen, S.F., Mátiás, G., Munro, R.C., Popescu, F., Slade, D., Tabbener, H., de Vries, S.M.G., Ziegenhagen, B., de Beaulieu, J.-L., Kremer, A., 2002b. Identification of refugia and post-glacial colonization routes of European white oaks based on chloroplast DNA and fossil pollen evidence. For. Ecol. Manage. 156, 49–74.

Petit, R.J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A., Kremer, A., 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. Proc. Natl. Acad. Sci. USA 94, 9996–10001.

Pignatti, S., 1982. Flora d'Italia. Edagricole, Bologna, Italy.

Posada, D., Crandall, A., 1988. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Rauscher, J.T., Doyle, J.J., Brown, H.D., 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the Glycine tomentella (Leguminosae) polyploid complex. Mol. Ecol. 11, 2691–2702.

Samuel, R., Bachmair, A., Jobst, J., Ehrendorfer, F., 1998. ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asiatic and North American taxa of *Quercus* (*Fagaceae*). Plant Syst. Evol. 211, 129–139.

Sauvage, C.H., 1961. Recherches géobotaniques sur les subérales marocaines. Série Botanique b, 21. I.S.C. Rabat.

Schirone, B., Schirone, A., Romagnoli, M., Angelaccio, C., Bellarosa, R., 1990. Preliminary considerations on the taxonomy of *Quercus crenata* Lamk. In: Lorenzoni, G.G., Ruggiero, L., Valenziano, S. (Eds.), Approcci metodologici per la definizione dell'ambiente fisico e biologico Mediterraneo, pp. 423–452 (in Italian).

Schirone, B., Spada, F., 1995. Anomalies in reproductive phenology and vegetation history: the case of SE Italy. Coll. Phytosoc. 14, 847–857.

Schirone, B., Spada, F., 2001. Some remarks on the conservation of genetic resources of Mediterranean oaks. In: Mediterranean Oaks Network, 1st Meeting, 12–14 October 2000, Antalya, Turkey IPGRI, Rome, pp. 21–26.

Schwarz, O., 1936–1939. Monographie der Eichen Europas, besonders des Mittelmeergebietes. Feddes Rep., Sonderbeih. D, 1–200.

Schwarz, O., 1964. *Quercus* L.. In: Tutin, T.G., Heywood, V.H., Burges, N.A., Morre, D.M., Valentine, D.H., Walters, S.M., Webb, D.A. (Eds.), Flora Europaea 1, 1st ed. Cambridge University Press, Cambridge, pp. 72–76.

Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic analyses. Syst. Biol. 49, 369–381.

Spada, F., Valfrè, D., Conti, F., Isopi, R., 1996. Long term dynamics in C Apennines (peninsular Italy) I: the case of *Betula pendula* Roth. Coll. Phytosoc. 24, 859–871.

Swofford, D.L., 1998. PAUP4.0b1*. Phylogenetic Analysis Using Parsimony. Test version. Sinauer Associates, Sunderland.

Taberlet, P., Fumagalli, L., Wust-Saucy, A.G., Cosson, J.F., 1998. Comparative phylogeography and postglacial colonization routes in Europe. Mol. Ecol. 7, 453–464.

Templeton, A.R., 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37, 221–244.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Toumi, L., Lumaret, R., 2001. Allozyme characterisation of four Mediterranean evergreen oak species. Biochem. Syst. Ecol. 29, 799–817.

Trelease, W., 1924. The American oaks. Mem. Natl. Acad. Sci 20, 1–255.

Van Valen, L., 1976. Ecological species, multispecies and oaks. Taxon 25, 233–239.

Whittemore, A.T., Schall, B.A., 1991. Interspecific gene flow in sympatric oaks. Proc. Natl. Acad. Sci. USA 88, 2540–2544.

Young, N.D., Healy, J., 2003. GapCoder automates the use of indel characters in phylogenetic analysis. BMC Bioinformatics 4, 6.

Zoldos, V., Siljak-Yakovlev, S., Papes, D., Sarr, A., Panaud, O., 2001. Representational difference analysis reveals genomic differences between *Q. robur* and *Q. suber*: implications for the study of genome evolution in the genus *Quercus*. Mol. Genet Genom. 265, 234–241.

Zuker, M., 1989. On finding all suboptimal foldings of an RNA molecule. Science 244, 48–52.

Zuker, M., Jacobson, A.B., 1998. Using reliability information to annotate RNA secondary structures. RNA 4, 669–679.

Zuker, M., Mathews, D.H., Turner, D.H., 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: Barciszewski, J., Clark, B.F.C. (Eds.), RNA Biochemistry and Biotechnology. NATO ASI Series. Kluwer Academic Publishers, Dordrecht, pp. 11–43.