Theor Appl Genet (1997) 94:947–955

P. Lashermes · M. C. Combes
P. Trouslot · A. Charrier

Phylogenetic relationships of coffee-tree species (*Coffea* L.) as inferred from ITS sequences of nuclear ribosomal DNA

Received: 25 July 1996 / Accepted: 18 October 1996

Abstract Phylogenetic relationships of *Coffea* species were estimated from the sequences of the internal transcribed spacer (ITS 2) region of nuclear ribosomal DNA. The ITS 2 region of 37 accessions belonging to 26 *Coffea* taxa and to three *Psilanthus* species was directly sequenced from polymerase chain reaction (PCR)-amplified DNA fragments. The level of variation was high enough to make the ITS 2 a useful tool for phylogenetic reconstruction. However, an unusual level of intraspecific variation was observed leading to some difficulty in interpreting rDNA sequence divergences. Sequences were analysed using Wagner parsimony as well as the neighbour-joining distance method. *Coffea* taxa were divided into several major groups which present a strong geographical correspondence (i.e. Madagascar, East Africa, Central Africa and West Africa). This organisation is well supported by cytogenetic evidence. On the other hand, the results were in contradiction with the present classification of coffee-tree taxa into two genera, namely *Coffea* and *Psilanthus*. Furthermore, additivity of parental rDNA types was not observed in the allotetraploid species *C. arabica*.

Key words *Coffea* · Coffee-tree · Internal transcribed spacer region · Nuclear ribosomal DNA · Molecular phylogeny

Introduction

Two genera, *Coffea* and *Psilanthus*, are distinguished in the *Coffeeae* tribe based on flowering and flower characteristics (Leroy 1980; Bridson 1987). All *Coffea* species are native to the inter-tropical forest of Africa and Madagascar, while species belonging to the genus *Psilanthus* originate from either Asia or Africa. Each genus has been divided into two subgenera (Bridson and Verdcourt 1988).

Coffee-trees differ greatly in morphology, size and ecological adaptation, thereby leading to the description of a large number of species. Particular attention has been paid to the subgenus *Coffea* (genus *Coffea* L.) which includes two cultivated species of economic importance, *Coffea arabica* L. and *Coffea canephora* Pierre (Berthaud and Charrier 1988). *C. arabica* is tetraploid (2n = 4x = 44) and is self-fertile while other *Coffea* species are diploid (2n = 2x = 22) and generally self-incompatible. Approximately 100 *Coffea* taxa have so far been characterised. Nevertheless, *Coffea* species hybridise readily with one another and produce relatively fertile hybrids (Charrier 1978; Louarn 1992). Infrageneric classifications have been proposed based on morphological characters (Lebrun 1941; Chevalier 1947). However, grouping criteria have become very complex and rather confused, and to-date they are considered of low value (Bridson and Verdcourt 1988). Complementary investigations are therefore required to clarify the phylogenetic relationships among these taxa (Charrier and Berthaud 1985). Biochemical components of coffee beans have been investigated (Clifford et al. 1989; Rakotomalala 1992; Anthony et al. 1993). However, such characteristics represent functional information and can be phylogenetically misleading due to parallel evolution and rapid adaptive radiation. Furthermore, genetic relationships among *Coffea* species have been assessed through molecular markers such as isozyme (Berthou and Trouslot 1977) and random amplified polymorphic DNA (RAPD, Lashermes et al. 1993). Although valuable results were obtained, the small number of species analysed and the nature of the molecular data did not allow a phylogenetic reconstruction. More recently, approaches

Communicated by P. M. A. Tigerstedt
P. Lashermes (✉) · M. C. Combes · P. Trouslot
ORSTOM, Laboratoire de ressources génétiques et d'amélioration des plantes tropicales, BP 5045, F-34032, Montpellier, France

A. Charrier
ENSAM, place Viala, F-34060, Montpellier, France

based on the chloroplast DNA (cpDNA) have been initiated but were handicapped by the low cpDNA variation observed in *Coffea* (Cros 1994; Lashermes et al. 1996 b).

In plants, the nuclear ribosomal DNA units (rDNA) consist of the 18*s*, 5.8*s* and 26*s* coding regions separated by intergenic spacers. Nuclear rDNA has proven to be a powerful phylogenetic tool because of the ubiquity of rDNA throughout plant species, the development of techniques for the rapid determination of the primary nucleotide sequence, and the diverse rates of evolution within and among component subunits and spacers (Hamby and Zimmer 1992). Whereas the 18*s* and 26*s* coding regions have been used to address phylogenetic questions at the family level or higher taxonomic levels, the internal transcribed spacers (ITS 1 and ITS 2) appear to be useful for assessing relationships at lower taxonomic levels because the sequences of spacer regions generally evolve more rapidly than the coding regions. Recently, the ITS region has been shown to be useful for resolving phylogenetic relationships among several plant genera, including *Antennaria* (Bayer et al. 1996), *Calycadenia* (Baldwin 1992, 1993) and *Krigia* (Kim and Jansen 1994).

In the present study, we have sequenced the internal transcribed spacer (ITS 2) region of 37 accessions representing 26 *Coffea* taxa and three *Psilanthus* species. The main purpose was to use the ITS sequences to attempt to resolve phylogenetic relationships among the closely related but highly diversified taxa associated with the genus *Coffea*. Additionally, we intended to evaluate the degree of divergence between both *Coffea* and *Psilanthus* genera.

Materials and methods

Plant samples

The names and origins of the 37 accessions selected for this study are listed in Table 1. The plant material was obtained from the ORSTOM collection resulting from several expeditions in Africa

Table 1 Origin of the accessions analysed for ITS 2 variation

Taxa	Accession code	Population name	Country of origin
1 *Coffea arabica* L.	ET 12-5		Ethiopia
2	Caturra		Brazil (cultivar)
3 *C. bertrandi* Chev.	Bertrandi		Madagascar
4 *C. brevipes* Hiern		Mt Cameroon	Cameroon
5 *C. canephora* Pierre	IF 444		Ivory Coast (cultivar)
6 *C. congensis* Froehner	03 255	Louma	Central African Rep.
7	03 1650	Brazzaville	Congo
8 *C. costatifructa* Bridson	08 111	Utete	Tanzania
9 *C. dolichophylla* Leroy	Dolichophylla		Madagascar
10 *C. eugenioides* Moore	04 1485	Cheptuyet	Kenya
11	04 010	Malava	Kenya
12	04 005	Kimilili	Kenya
13 *C. eugenioides* var. *kiwuensis* (Lebrun) Chev.	Kiwuensis		Uganda
14 *C. farafanganensis* Leroy	Farafanganensis		Madagascar
15 *C. humilis* Chev.	07 141	Sakré	Ivory Coast
16 *C. kapakata* Chev.	intro. Brazil		Angola
17 *C. liberica* Hiern	EC 16	Koto	Cameroon
18 *C. liberica* var. *dewevrei* Lebrun	05 797	N'Dongue	Central African Rep.
19 *C. liberica* var. *liberica* (Hiern) Lebrun	05 242	Taï	Ivory Coast
20 *C. millotii* Leroy	Millotii		Madagascar
21 *C. perrieri* Drake	Perrieri		Madagascar
22 *C. pseudozanguebariae* Bridson	08 021		Kenya
23 *C. racemosa* Lour.	intro. Portugal		Mozambique
24 *C. resinosa* (Hook.) Radlk.	Resinosa		Madagascar
25 *C. salvatrix* Swynn. & Phil.	intro. Brazil		Mozambique
26 *C. sakarahae* Leroy	Sakarahae		Madagascar
27 *C. sessiliflora* Bridson	PB 70	Kitulangalo	Tanzania
28 *C. sp.* Mayombe		Mayombe	Congo
29 *C. sp.* Moloundou	OC 210	Souanké	Congo
30 *C. sp.* Moloundou	OC 204	Moloundou 1	Cameroon
31 *C. sp.* N'gongo II		N'gongo II	Congo
32 *C. sp.* N'koumbala	OC 105	N'koumbala	Cameroon
33 *C. sp.* X			unknown
34 *C. stenophylla* Don	FB 1	Ira	Ivory Coast
35 *Psilanthus ebracteolatus* Hiern	OA 153		Ivory Coast
36 *P. mannii* Hook. f.	OA 009		Ivory Coast
37 *P. travancorensis* (Wight & Arn.) Leroy	Travancorensis		India

and Madagascar (Anthony 1992). Thirty-four of the accessions belonged to 26 *Coffea* taxa. Most of the taxa available were studied. Nevertheless, with the species native to Madagascar, only three of the six described botanical series (reviewed in Charrier 1978) were represented by one to three species. In addition to *Coffea* taxa, three species of the closely related genus *Psilanthus* were represented by one accession: *P. mannii*, of the subgenus *Psilanthus*, *P. ebracteolatus* and *P. travancorensis*, which are included in the subgenus *Afrocoffea* (Bridson and Verdcourt 1988).

Total DNA was extracted from lyophilised leaves through a nuclei isolation step as described by Paillard et al. (1996) with slight modifications. In particular, the nuclear lysis solution was replaced by a buffer containing 0.1 M Tris-HCl (pH 8.0), 0.02 M EDTA, 1.25 M NaCl and 4% MATAB (mixed alkyl tri-methyl ammonium bromide).

PCR amplification and DNA sequencing

The entire ITS region (ITS1-5.8*s*-ITS2) was amplified with primers ITSL (5'-TCGTAACAAGGTTTCCGTAGGTG-3'; Hsiao et al. 1994) and ITSR (5'-TATGCTTAAAYTCAGCGGG-3'; sequence provided by V. Savolainen). Primer ITSL anneals to 18*s* rDNA near the ITS 1 border, while ITSR is complementary to 25*s* rDNA near the ITS 2 border. Amplifications were performed in a 50-µl vol containing 10 mM Tris HCl, pH 9.0, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 50 mM KCl, 150 µM of each dATP, dCTP, dGTP, dTTP, 0.5 µM of each primer, 50 ng of total DNA, and 1 U of *Taq* polymerase (Promega). Reactions were performed in a PTC-100 thermal cycler (MJ Research). After 5 min heating at 95°C, 35 cycles were run. Each cycle consisted of 1 min at 95°C, 1 min at 58°C and 1 min at 72°C. This was followed by 4 min at 72°C. The amplification products were purified by agarose-gel electrophoresis, and the concentrated DNAs were recovered using fibreglass (Appligene, France).

Direct sequencing was done from the double-stranded DNA fragment with one of the amplification primers or one internal sequencing primer ITS2L (5'-CCRCGAACCATCGAGTCTTTG-3') which anneals to 5.8*s* rDNA near the ITS 2 border, using a dideoxy chain-termination reaction according to the Applied Biosystems autosequencing protocol. The PCR products were analysed on an ABI373A autosequencer.

Sequence analyses

Boundaries of the ITS 2 region were determined by comparison with known sequences for the 5.8*s* and 25*s* coding regions of nuclear rDNA (Takaiwa et al. 1985; Kavanagh and Timmis 1988; Baldwin 1992). Sequences obtained from the 37 coffee-tree accessions were aligned with the CLUSTAL V multiple-sequence alignment program (Higgins et al. 1992).

The aligned sequences were analysed by the neighbour-joining (NJ) tree construction method (Saitou and Nei 1987) using a TREECOM software package (Van de Peer and De Wachter 1994). Insertions and deletions (indels) were taken into account. The distance according to the Jukes and Cantor (1969) model was calculated as:

$$D_{AB} = -\frac{3}{4}\ln\left[1-\frac{4}{3}\left(\frac{S_U}{I+S_U}\right)\right]\left[1-\frac{G}{T}\right]+\frac{G}{T},$$

where D_{AB} is the distance between sequences A and B, I the number of identical nucleotides, S_u the number of positions showing a substitution, G the number of gaps in one sequence with respect to the other, and T the sum of I, S and G. Cladistic analyses were also performed. Only aligned nucleotide sites with potential phylogenetic information, i.e. with each of at least two nucleotide states in two or more sequences, were included in the data matrix. Indels were also included as single characters. Wagner parsimony phylogenetic trees (Farris 1970) were constructed with the phylogenetic inference package (PHYLIP, version 3.4) written by Felsenstein (1989). The DNAPARS program was used to find the most-parsimonious trees. Shortest parsimonious trees were used to construct a strict consensus tree using the program CONSENSE. The bootstrap method (Felsenstein 1985) was employed to evaluate the reliability of tree topologies.

Results

Organisation and overall sequence variation of the ITS region

The aligned sequences of the ITS 2 region, the 5.8*s* subunit and the extremity of the ITS 1 region of *C. canephora* and *C. millotii*, along with the published sequences of melon (*Cucumis melo*, Kavanagh and Timmis 1988) and rice (Takaiwa et al. 1985), are given in Fig. 1. The configuration of the entire ITS region of both *Coffea* species was similar to that of other plants. The region encoding the 5.8*s* rDNA showed very high similarity among all four species. Sequence divergence from pairwise comparisons between the two *Coffea* species and either *C. melo* or *Oryza sativa* was 5%. The 5.8*s* subunit was uniform in size (164 bp) among the nucleotide sequences of both *Coffea* species and only one variable site was observed. On the other hand, the ITS 2 region showed significant sequence variation among the different genera as well as between the two *Coffea* species. Mean sequence divergence between the two *Coffea* species and either *C. melo* or *O. sativa* was 44% and 50%, respectively. Comparisons of *C. canephora* and *C. millotii* ITS 2 sequences indicated nine variable positions (4%) including five transitions, three transversions and one short indel. On the basis of these preliminary results, only the ITS 2 sequences were established for the remaining coffee-tree accessions.

ITS 2 sequence divergence

The sequence data have been deposited in the EMBL/GenBank/DDBJ nucleotide sequence databases under the accessions numbers U63811 and U64320 to U64355. The lengths of the ITS 2 sequences of the 34 accessions of *Coffea* and the three *Psilanthus* samples analysed ranged from 199 bp in *P. ebracteolatus* to 211 bp in *C. liberica*. These variations in length were attributable to deletion and insertion events, and gaps were introduced to align the sequences. Alignment of all *Coffea* and *Psilanthus* ITS 2 sequences resulted in 234 characters and necessitated 23 gaps. With the notable exception of five indels (4–8 bp), the gaps were 1 bp in length. Gaps were correlated with particular species groups and were of potential value for phylogenetic reconstruction.

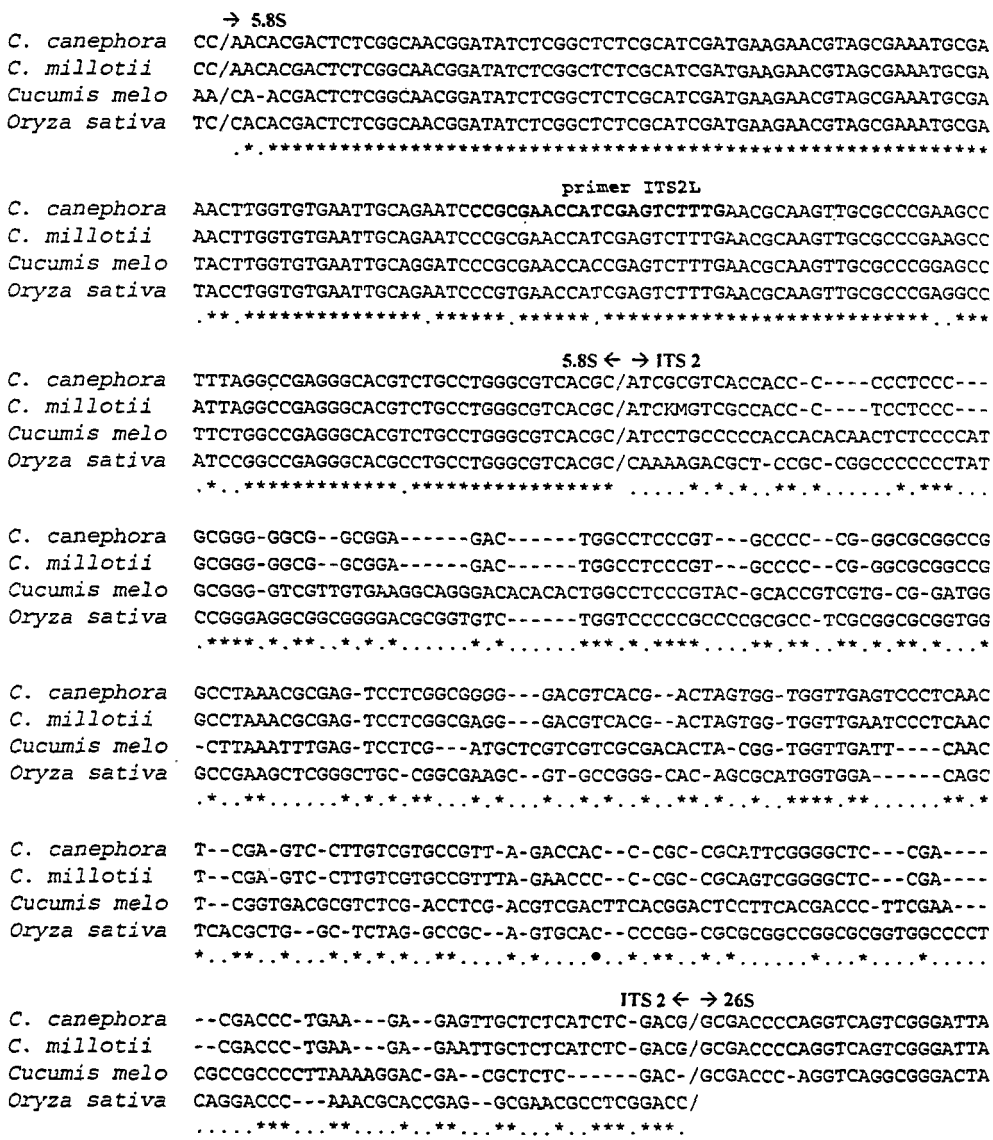

Fig. 1 Aligned nucleotide sequence of the 5.8*s* and ITS 2 regions of two *Coffea* species (*C. canephora* and *C. millotii*), *Cucumis melo* and *Oryza sativa*. *Arrows* indicate boundaries of the different regions; *dashes* denote gaps; * denote identity of the four sequences. The position of the internal sequencing primer (ITS2L) is also indicated

The pairwise nucleotide-sequence divergence (Jukes and Cantor one-parameter distance) among the *Coffea* taxa (Table 2) ranged from 1.5% between *C. eugenioides* and *C. sp.* Moloundou to 39% between *C. racemosa* and *C. sakarahae*. Pairwise comparisons between the *Coffea* taxa and the *Psilanthus* accessions indicated sequence divergences ranging from 13% to 34%. Average divergence of the *Psilanthus* species, *P. ebracteolatus*, *P. mannii* and *P. travancorensis*, from the *Coffea* taxa were 18%, 24% and 18%, respectively. Significant variation among accessions of the same species from distant geographical regions was detected. Infraspecific sequence divergence ranged from 3.0% in *C. sp.* Moloundou and *C. arabica* to 8.3% in *C. congensis*. In addition, individual DNA sequences exhibited polymorphic nucleotide sites that could indicate multiple rDNA repeat types.

Phylogenetic inference

The neighbour-joining tree obtained using the sequence divergences calculated by the Jukes and Cantor one-parameter method (Table 2) is shown in Fig. 2. Maximum-parsimony analysis yielded 100 maximally parsimonious trees. Each required 174 evolutionary steps (consistency index = 0.66). The strict consensus tree is presented in Fig. 3. The tree was rooted by *P. mannii* because this species showed steady divergence from all *Coffea* taxa. The topologies of the phylogenetic trees generated by these two reconstruction methods were roughly congruent. As expected, association of accessions belonging to the same species appeared stronger with the Wagner-parsimony method applied to the data matrix including only the character with potential phylogenetic information. *C. arabica*

Table 2 Pairwise comparisons of ITS 2 sequence obtained from the 37 coffee-tree accessions listed in Table 1. The percentage of divergence is estimated according to the Jukes and Cantor model

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	3.0																		
3	15.6	13.7																	
4	4.0	2.5	13.8																
5	2.0	5.0	14.9	4.0															
6	4.0	2.0	14.3	2.5	4.5														
7	12.3	10.8	18.9	9.3	13.4	8.3													
8	13.7	15.7	19.2	14.5	13.1	15.0	20.8												
9	18.5	16.8	12.0	16.9	17.9	16.2	18.1	20.5											
10	12.9	11.3	15.8	9.3	11.8	10.4	16.8	16.4	19.3										
11	5.6	8.3	15.0	7.3	5.1	7.8	15.9	10.9	16.6	7.6									
12	11.8	11.3	17.7	10.3	11.2	10.2	17.7	13.4	17.8	7.4	5.6								
13	9.7	9.2	14.7	7.7	9.1	8.2	15.2	13.5	18.3	4.5	5.0	6.4							
14	13.3	14.0	13.8	14.1	12.2	13.5	24.8	18.2	19.5	18.2	12.5	15.0	16.0						
15	7.8	9.5	15.1	8.4	7.8	8.9	15.4	12.9	17.0	12.5	7.3	10.3	9.3	14.2					
16	6.1	6.6	15.6	5.1	5.5	6.6	12.9	14.5	16.6	12.9	7.7	11.2	10.8	15.1	10.6				
17	13.3	13.4	16.4	13.5	12.7	12.8	20.7	16.9	17.4	16.5	11.7	14.1	14.9	18.1	14.2	13.9			
18	6.6	8.2	11.8	7.2	6.0	7.7	15.7	11.3	16.5	11.2	5.1	9.6	8.6	12.2	7.2	8.2	8.2		
19	8.7	10.5	13.6	9.4	8.2	9.9	17.0	13.8	14.8	13.5	7.8	13.5	12.5	15.1	10.0	9.8	6.7	4.5	
20	5.5	8.8	11.7	6.6	4.5	8.2	16.3	12.5	15.0	11.2	4.1	9.7	8.6	9.1	6.7	7.6	12.1	5.5	8.2
21	10.7	11.3	6.1	9.0	9.5	11.3	15.6	15.0	12.6	11.5	10.2	13.3	11.1	13.8	12.1	9.7	16.2	10.1	11.3
22	15.2	14.7	20.5	13.2	14.6	13.7	21.2	10.9	23.4	15.1	11.2	13.7	12.3	20.1	13.9	14.7	18.4	11.6	15.9
23	16.4	20.2	28.2	18.7	17.0	19.2	27.0	16.5	30.3	21.8	14.9	20.8	19.5	24.7	19.4	20.7	23.4	18.8	20.7
24	20.9	19.8	11.5	18.6	19.6	21.1	27.0	24.9	19.6	21.6	18.6	22.3	21.2	19.0	20.5	18.0	21.1	17.0	17.7
25	16.9	20.9	26.8	20.3	16.3	19.9	25.9	15.1	27.0	20.2	14.7	20.2	18.5	24.6	18.8	20.1	23.6	16.9	18.9
26	24.9	21.9	13.3	20.7	23.6	21.2	25.5	28.6	16.6	21.8	23.9	23.1	22.0	26.0	24.0	22.8	25.9	22.9	24.4
27	9.9	13.5	19.8	11.7	9.4	12.8	20.2	5.8	21.4	14.2	6.7	11.9	11.5	17.1	9.7	12.2	15.3	9.4	11.7
28	11.4	12.1	15.1	10.4	10.8	11.5	19.2	15.0	17.7	13.9	11.0	13.3	12.3	15.5	11.0	11.0	14.4	8.1	11.4
29	7.2	6.2	13.2	6.2	6.1	6.1	15.3	10.9	16.8	5.4	1.5	6.0	3.4	12.9	7.8	8.3	11.1	5.6	8.8
30	10.2	8.1	13.5	6.2	9.1	8.3	15.2	14.1	17.7	5.4	4.5	7.4	2.4	16.0	9.9	10.3	14.9	8.1	11.4
31	7.8	7.3	11.7	6.8	7.8	6.7	13.1	12.3	15.1	10.9	8.4	10.9	8.7	14.1	9.0	6.8	12.4	5.6	9.5
32	14.1	10.8	14.6	8.8	14.6	9.3	12.8	20.2	18.8	14.2	16.0	16.3	13.8	19.3	16.1	14.7	18.4	13.5	16.5
33	8.1	6.6	14.2	7.1	9.2	7.2	13.6	13.6	19.0	8.5	5.0	9.0	6.4	15.5	9.9	9.8	14.4	8.1	10.9
34	9.2	8.7	13.2	7.8	9.2	8.8	17.7	13.2	15.3	11.9	8.2	11.7	10.3	13.8	8.8	8.8	12.1	6.5	8.1
35	16.3	18.4	22.6	16.2	15.7	17.3	23.8	15.7	26.3	16.2	13.5	15.9	13.4	19.5	13.2	17.6	22.9	16.3	18.9
36	23.0	20.0	27.2	17.8	22.4	20.2	27.4	23.3	32.3	22.0	20.6	22.2	20.3	27.8	21.0	22.3	28.8	23.0	25.1
37	16.0	15.0	16.9	13.1	16.0	14.7	20.1	21.6	20.8	16.5	15.7	17.4	13.8	21.9	16.4	15.3	21.0	15.4	18.5

	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
21	6.7																
22	14.0	17.6															
23	18.2	22.6	15.8														
24	15.8	11.6	24.7	32.9													
25	16.9	20.7	15.5	20.3	33.4												
26	21.0	10.8	31.7	39.0	22.6	37.2											
27	8.8	14.4	8.9	13.1	24.6	1.7	29.6										
28	10.8	11.3	15.9	23.0	20.4	20.3	23.2	14.4									
29	5.6	10.2	10.1	16.6	18.5	16.5	21.9	8.3	10.9								
30	8.6	10.0	12.3	19.5	19.3	18.6	20.8	11.5	11.8	3.0							
31	8.3	9.3	14.1	21.7	17.7	19.8	20.4	13.0	5.6	7.3	8.8						
32	16.4	13.3	19.7	27.2	24.4	26.0	22.4	20.3	14.7	14.2	14.1	10.3					
33	8.6	11.7	13.9	20.5	20.7	19.6	24.1	12.1	11.3	4.0	4.9	7.7	14.0				
34	8.6	9.6	13.5	20.9	17.8	20.5	21.2	12.1	4.5	8.2	8.7	5.6	14.2	9.1			
35	15.1	16.2	13.8	20.1	25.2	20.3	30.3	12.9	18.0	13.5	14.6	14.8	21.3	16.2	15.9		
36	23.0	21.7	19.6	27.8	31.4	28.8	34.1	19.4	25.1	19.4	19.3	22.6	26.0	21.6	21.3	16.0	
37	13.2	13.2	20.4	27.7	25.4	26.4	22.3	19.4	17.9	15.0	14.4	13.6	18.4	17.7	15.2	21.0	25.8

exhibited only one ITS 2 sequence type, which for both accessions analysed appeared very similar to the sequences of canephoroid species (*C. canephora*, *C. congensis*, *C. brevipes*).

A number of groups of *Coffea* taxa were consistently obtained in both trees, although the bootstrap values were heterogeneous and moderate on average. Geographical distribution of taxa included in the major

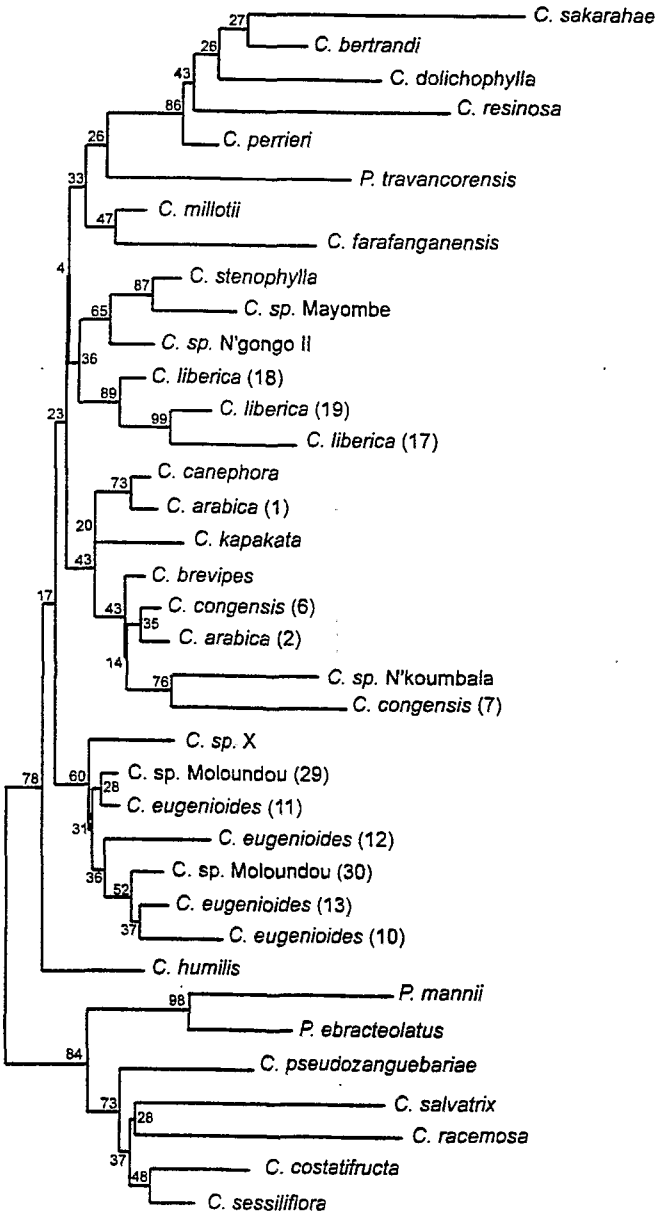

Fig. 2 Neighbour-joining tree of 37 coffee-tree accessions using a Jukes and Cantor one-parameter distance. *Numbers* on the branches are bootstrap values (%) obtained from 200 replicate analyses

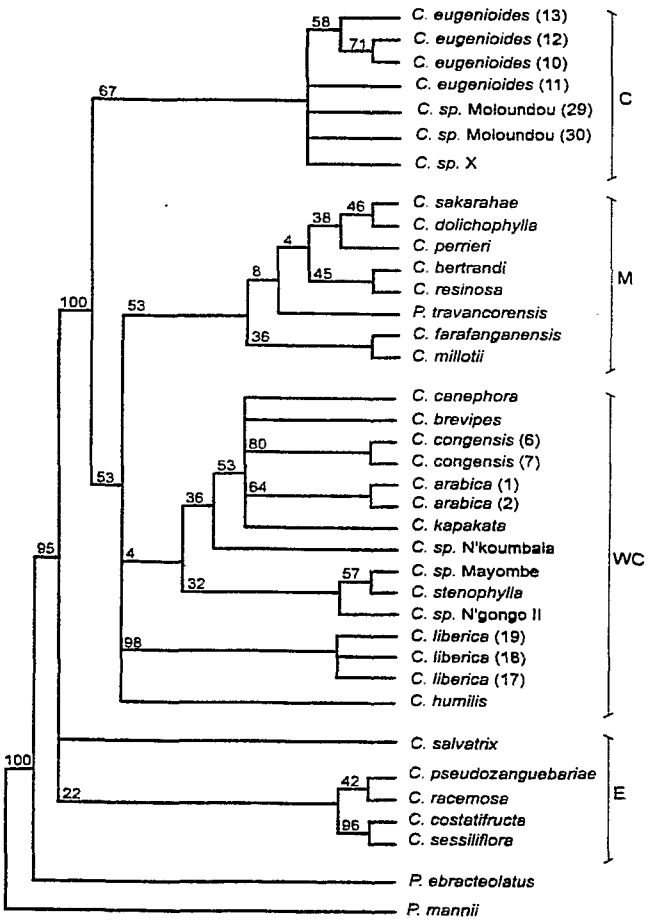

Fig. 3 Strict consensus Wagner tree, constructed from 100 equally most-parsimonious trees resulting from phylogenetic analysis of ITS 2 sequence variation data for coffee-tree taxa. *P. mannii* was used as the outgroup species. Frequenies of occurrence of a monophyletic group among 100 bootstrap replicate are shown above the line at each node. Major *Coffea* groups are indicated by *letters* according to the geographical origin of accessions: *C* (Central Africa), *M* (Madagascar), *WC* (West and Central Africa) and *E* (East Africa)

groups of taxa resulting from the Wagner-parsimony analysis is presented in Fig. 4. A clear correspondence was observed between the geographical origin of taxa and the phylogenetic grouping. The groups consisted of *Coffea* taxa originate from West and Central Africa, Central Africa, East Africa, and Madagascar, respectively. Overlapping geographical distributions of taxa-groups was noted only in the central part of Africa where two groups were distinguished. Neither phylogenetic analysis was clear on the presence of one unique or several clades in the group of taxa found in West and Central Africa. In particular, *C. humilis* seemed distantly related to other West-African taxa.

One unexpected result was that both parsimony and distance analyses placed *P. ebracteolatus* and *P. mannii* as the sister group to a clade consisting of East-African *Coffea* species. Moreover, *P. travancorensis* was placed with the *Coffea* species originating from Madagascar.

Discussion

Organisation of the nuclear rDNA repeat unit of the *Coffea* species seemed similar to that of other plants: a 5.8*s* coding region flanked by two internal transcribed spacers and located between the 18*s* and 26*s* coding regions. As predicted, the ITS 2 region appeared much

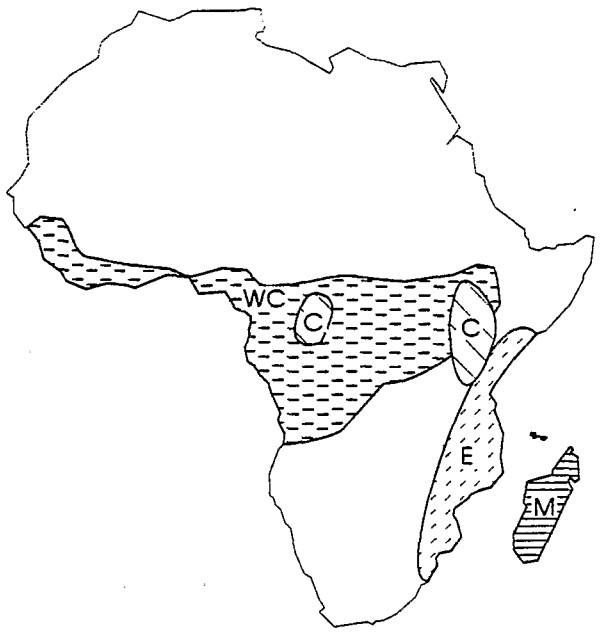

Fig. 4 Geographical distribution of the major groups of *Coffea* taxa revealed by the ITS 2 sequence-variation analysis (see Fig. 3)

more divergent than the 5.8*s* coding region. Even though the sequences of the internal transcribed spacers are expected to be conserved to some extent, because of their important role in post-transcriptional processing (Hamby and Zimmer 1992), the level of variation was high enough to make the ITS 2 a useful tool for phylogenetic reconstruction at the species level in *Coffea*. Therefore, these data provide additional support for the utility of the ITS region for phylogenetic investigation among closely related dicot species.

Intraspecific ITS variation

The overall sequence homogeneity among members of a gene family such as the nuclear rDNA is assumed to be maintained by homogenisation mechanisms associated with concerted evolution (Arnheim 1983). As a result, rDNA repeats are usually very similar within individuals and species, although differences may accumulate between species (Hillis and Dixon 1991). An unusual feature of the ITS 2 sequence in *Coffea* was the importance of intraspecific variation. Even the presence of ITS variants within individuals was observed. Similar intraspecific variation has already been reported. For instance, Baldwin (1993) found 4.3% intraspecific divergence in *Calycadenia truncata*. However, this phenomenon appeared particularly important within coffee-tree species. Such results are expected when the rate of nucleotide divergence exceeds the homogenisation rate of the gene copies within a multigene family, a situation that could arise in cases of explosive radiation and/or interspecific hybridisation (Hillis and Davis 1988).

The generation times of coffee-trees have been estimated as between 20 and 30 years (Berthaud 1986). The observed deficiency in the homogenisation mechanisms may be related to the long life cycles of coffee-trees, in the same manner that nucleotide-substitution rates have been reported to be related to the length of the reproductive cycle (Li and Graur 1991). Moreover, it is most likely that spontaneous interspecific hybridisation occurs between taxa and has been involved in speciation. The results of the present study on nuclear rDNA, as well as initial analysis of chloroplast DNA variation (Cros 1994), conformed to a radial mode of speciation for the coffee-tree species. In addition, a recent origin was suggested for the genus *Coffea* based on the low level of variation exhibited by chloroplast DNA (Lashermes et al. 1996b). Therefore, the high level of intraspecific variation observed for the ITS 2 region in this study most likely reflects biological characteristics and the evolutionary history of *Coffea* species.

This phenomenon constitutes a difficulty in interpreting rDNA sequence divergences. On the other hand, within-species variation in ITS sequences offers the opportunity for resolving relationships among distinct populations and for addressing questions of species boundaries. It is noteworthy that the sequence divergence among accessions of *C. congensis* was greater than the variation noted between *C. congensis* and either *C. brevipes* or *C. canephora* accessions. Since all three species are easily crossable and produce highly fertile hybrids (Louarn 1992), taxonomic criteria appear questionable. Comparable results were observed with *C. eugenioides* and *C. sp.* Moloundou, and it would be particularly interesting to perform interspecific hybridisation between these taxa.

The case of the polyploid *C. arabica*

C. arabica is considered to be a segmental allotetraploid species (Carvalho 1952; Grassias and Kammacher 1975). Recently, the origin of *C. arabica* was corroborated and specified using DNA-based markers (Lashermes et al. 1996a). Biparental inheritance and additivity of parental rDNA types have been documented in several hybrid plant taxa (Doyle et al. 1985; Doyle and Doyle 1988). In the case of *C. arabica*, the presence of only one type of ITS 2 sequence is most likely due to the homogenising effect of concerted evolution and/or possible backcrossing with parental species. Furthermore, the present results showed that a species related to the canephoroid group (*C. canephora*, *C. congensis*, *C. brevipes*) was one of the progenitor species.

ITS sequence phylogeny of the *Coffea* species

The phylogenetic relationships of *Coffea* taxa inferred from ITS 2 sequences indicated four major groups with a strong geographical correspondence. One group included all species native to Madagascar (*Mascarocoffea*). The large variation observed in this group reflects the considerable diversity of coffee-trees found in Madagascar (Charrier 1978). Species endemic to the region between the Kivu ridge and the Mozambique Channel (*Mozambicoffea*) formed a second group. *C. eugenioides*, which is native to the uplands of ridge region, was classified with two taxa, *C. sp.* Moloundou and *C. sp.* X. which are the only two diploid taxa reported to be self-fertile. While the origin of *C. sp.* X is unknown, *C. sp.* Moloundou was recently discovered in the Congo basin (Anthony 1992). The small number of taxa included in this group could either be an artefact due to the lack of substantial collecting missions on the west side of the Kivu ridge, or the consequence of specific biological characteristics. The last group encompasses diploid species originating from West and Central Africa, and the allotetraploid *C. arabica*. This group appeared heterogeneous and subgroups could be distinguished such as the canephoroid group. Moreover. *C. humilis* could even be considered as the unique representive of an additional major group. The relatively complex patterns of distribution of *Coffea* species in Central and West Africa may be related in part to glaciation phases during the quaternary period (Hamilton 1976) as suggested by Berthaud (1986).

These biogeographical groups are well supported by the cytogenetic data (Charrier 1978; Louarn 1992). Hybrids resulting from interspecific hybridisation between species of the same group are characterised by a high degree of bivalent meiotic chromosome associations and high fertility. In contrast, hybrids between species of different ITS-based groups display low fertility.

Within each biogeographical group, the relatively close ITS relationship of taxa suggests a common origin, with subsequent ecological differentiation leading to the considerable morphological variation observed. Cases of hybridisation between these interfertile entities could not be discounted, and could have substantial impact on phylogenetic estimation from ITS sequences. Additional data are therefore required to overcome limitations of the ITS region and to bring into better focus overall *Coffea* species relationships. Comparisons between phylogenies inferred from both nuclear and chloroplast genomes would be particularly interesting. Since the occurrence of interspecific hybridisation and plant introgression (Rieseberg and Brunsfeld 1992) may be detected.

Intergeneric relationships

Analysis of the ITS 2 sequences does not support the present division in the tribe *Coffea*, namely *Coffea* and *Psilanthus* (Leroy 1980; Bridson 1987). In comparison to the variation observed in the genus *Coffea*, the intergeneric divergence appeared relatively limited. Similar results have been observed with chloroplast DNA (Lashermes et al. 1996b) and recently intergeneric hybridisation has even been achieved (Couturon, personal communication). In contradiction with the present division of the genus *Psilanthus* into two subgenera, *P. travancorensis* (subgenus *Afrocoffea*) appeared distantly related to the two other *Psilanthus* species analysed, *P. ebracteolatus* (subgenus *Afrocoffea*) and *P. mannii* (subgenus *Psilanthus*). It is noteworthy that *P. travancorensis*, which is native to India, was placed with the *Coffea* species from Madagascar, while *P. ebracteolatus* and *P. mannii*, which are both present in Africa, showed similarity with *Coffea* species from East Africa. These results prompt modifications in the adopted classification and support the hypothesis of a common origin of coffee-trees in Africa.

Acknowledgements This work was supported in part by the European Community through the International Scientific Co-operation Programme (Contract CI1-CT91-0899).

References

Anthony F (1992) Les ressources génétiques des caféiers: collecte, gestion d'un conservatoire et évaluation de la diversité génétique. Collection TDM (81), ORSTOM (ed). Paris

Anthony F, Clifford MN, Noirot M (1993) Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. Genet Res Crop Evol 40:61–70

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer Assoc, Sunderland, pp 38–61

Baldwin BG (1992) Phylogenetic utility of the internal transcribed spacers of nuclear ribosomal DNA in plants: an example from the Compositae. Mol Phylogenet Evol 1:3–16

Baldwin BG (1993) Molecular phylogenetics of *Calycadenia* based on ITS sequences of nuclear ribosomal DNA: chromosomal and morphological evolution re-examined. Am J Bot 80:222–238

Bayer RJ, Soltis DE, Soltis PS (1996) Phylogenetic inferences in *Antennaria* based on sequences from nuclear ribosomal DNA internal transcribed spacers (ITS). Am J Bot 83:516–527

Berthaud J (1986) Les ressources génétiques pour l'amélioration des caféiers africains diploïdes. Collection "Travaux et Documents" (188), ORSTOM (ed), Paris

Berthaud J, Charrier A (1988) Genetics resources of *Coffea*. In: Clarke RJ, Macrae R (eds) Coffee, vol 4: agronomy. Elsevier Applied Science, London, pp 1–42

Berthou F, Trouslot P (1977) L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une méthode d'électrophorèse en série. 8th Int Sci Colloqium on coffee. ASIC, Paris, pp 373–383

Bridson DM (1987) Nomenclatural notes on *Psilanthus*, including *Coffea* sect. *Paracoffea* (*Rubiaceae* tribe *Coffeeae*). Kew Bull 42:453–460

Bridson DM, Verdcourt B (1988) Flora of tropical East Africa – Rubiaceae (part 2). Balkema, Brookfield, Rotterdam

Carvalho A (1952) Taxonomia de *Coffea arabica* L., Caracteres morfologicos dos haploides. Bragantia 12:201–212

Charrier A (1978) La structure génétique des caféiers spontanés de la région malgache (*Mascarocoffea*). Mémoires ORSTOM (87), ORSTOM (ed), Paris

Charrier A, Berthaud J (1985) Botanical classification of coffee. In: Clifford MN, Wilson KC (eds), Coffee: botany, biochemistry and production of beans and beverage. Croom Helm, London, pp 13–47

Chevalier A (1947) Les caféiers du Globe III. Systématique des caféiers et faux caféiers. Maladies et insectes nuisibles. Encyclopédie biologique 28 (3), Lechevalier, Paris

Clifford MN, Williams T, Bridson D (1989) Chlorogenic acids and caffeine as possible taxonomic criteria in *Coffea* and *Psilanthus*. Phytochemistry 30:4039–4040

Cros J (1994) Implications phylogénétiques des variations de l'ADN chloroplastique chez les caféiers (genres *Coffea* L. et *Psilanthus* Hook. f.). Doctoral thesis, Université Montpellier II, France

Doyle JJ, Doyle JL (1988) Natural interspecific hybridization in eastern North American *Claytona*. Am J Bot 75:1238–1246

Doyle JJ, Soltis DE, Soltis PS (1985) An intergeneric hybrid in the Saxifragaceae: evidence from ribosomal RNA genes. Am J Bot 72:1388–1391

Farris JS (1970) Methods for computing Wagner trees. Systematic Zool 19:83–92

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

Felsenstein J (1989) PHYLIP-phylogeny inference package. Cladistics 5:164–166

Grassias M, Kammacher P (1975) Observations sur la conjugaison chromosomique de *Coffea arabica* L.. Café Cacao Thé 19: 177–190

Hamby RK, Zimmer EA (1992) Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 50–91

Hamilton AC (1976) The significance of patterns of distribution shown by forest plants and animals in Tropica Africa for the reconstruction of Upper Pleistocene paleoenvironments: a review. Palaeoecology of Africa 9:63–97

Higgins DG, Bleasby AJ, Fuchs R (1992) CLUSTAL V: Improved software for multiple sequence alignment. CABIOS 8:189–191

Hillis DM, Davis SK (1988) Ribosomal DNA: intraspecific polymorphism, concerted evolution and phylogeny reconstruction. Systematic Zool 37:63–66

Hillis DM, Dixon MT (1991) Ribosomal DNA: molecular evolution and phylogenetic inference. Quart Rev Biol 66:411–453

Hsiao C, Chatterton NJ, Asay KH, Jensen KB (1994) Phylogenetic relationships of 10 grass species: an assessment of phylogenetic utility of the internal transcribed spacer region in nuclear ribosomal DNA in monocots. Genome 37:112–120

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Kavanagh TA, Timmis JN (1988) Structure of melon rDNA and nucleotide sequence of the 17–25*s* spacer region. Theor Appl Genet 76:673–680

Kim KJ, Jansen RK (1994) Comparisons of phylogenetic hypotheses among different data sets in dwarf dandelions (*Krigia*, Asteraceae): additional information from internal transcribed spacer sequences of nuclear ribosomal DNA. Plant Syst Evol 190:157–185

Lashermes P, Cros J, Marmey P, Charrier A (1993) Use of random amplified DNA markers to analyze genetic variability and relationships of *Coffea* species. Genet Res Crop Evol 40:91–99

Lashermes P, Combes MC, Trouslot P, Anthony F, Charrier A (1996a) Molecular analysis of the origin and genetic diversity of *Coffea arabica* L.: implications for coffee improvement. Proc EUCARPIA meeting on tropical plants, Montpeffier, pp 23–29

Lashermes P, Cros J, Combes MC, Trouslot P, Anthony F, Hamon S, Charrier A (1996b) Inheritance and restriction fragment length polymorphism of chloroplast DNA in the genus *Coffea* L. Theor Appl Genet 93:626–632

Lebrun J (1941) Recherches morphologiques et systématiques sur les caféiers du Congo. Mémoires TX1 (3), Institute Royal Colonial Belge, Bruxelles

Leroy JF (1980) Evolution et taxogenèse chez les caféiers: hypothèse sur l'origine. C R Acad Sci (Paris) 291:593–596

Li WC, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland, Massachusetts

Louarn J (1992) La fertilité des hybrides interspécifiques et les relations génomiques entre caféiers diploïdes d'origine africaine (Genre *Coffea* L. sous genre *Coffea*). Thesis, Université d'Orsay, Paris XI, France

Paillard M, Lashermes P, Pétiard V (1996) Construction of a molecular linkage map in coffee. Theor Appl Genet 93:41–47

Rakotomalala JJ (1992) Diversité biochimique des caféiers. Collection TDM (99), ORSTOM (ed), Paris

Rieseberg LH, Brunsfeld SJ (1992). Molecular evidence and plant introgression. In: Soltis PS, Soltis DE, Doyle JJ (eds) Molecular systematics of plants, Chapman and Hall, New York, pp 151–176

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Takaiwa F, Oono K, Sugiura M (1985) Nucleotide sequence of the 17*s*-25*s* spacer region from rice rDNA. Plant Mol Biol 4:355–364

Van de Peer Y, De Wachter R (1994) TREECOM for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput Applic Biosci 10:569–570