

## Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L.

F. Berthou

Biologie et Amélioration des Plantes Utiles, ORSTOM, Bondy, France

C. Mathieu and F. Vedel

Laboratoire de Photosynthèse, Biologie Cellulaire Végétale, F-91190 Gif-sur-Yvette, France

Received November 19, 1982; Accepted January 23, 1983

Communicated by D. von Wettstein

**Summary.** Chloroplast and mitochondrial DNA from nine species or taxons of coffee-trees were compared as to their phylogenetic relationship by restriction endonuclease fragment analysis. Three types of chloroplast DNA (cp DNA) were detected indicating relationships as follows: (i) *C. arabica*, *C. eugenioides*; (ii) *C. canephora*, *C. congensis*, "nana" taxon; (iii) *C. liberica*. The mitochondrial DNA (mt DNA) separated into five types: (i) *C. arabica*, *C. eugenioides*, *C. congensis*; (ii) *C. canephora*, "nana" taxon; (iii) *C. excelsa*; (iv) *C. liberica*; (v) *Paracoffea ebracteolata*. The divergence in organelle DNAs agrees with the phylogenetic relationship deduced by conventional methods and is presented in some detail. Restriction patterns of the cp and mt DNAs isolated from a clone of *C. arabusta* have been compared to those of the parents and were found to be inherited from the mother. Cp and mt DNA analyses in the genus *Coffea* support the hypothesis that *C. canephora* diverged from *C. congensis*, whereafter the latter species differentiated into *C. eugenioides* and *C. arabica*.

**Key words:** *Coffea* – Coffee-tree evolution – Chloroplast DNA – Mitochondrial DNA

### Introduction

The genus *Coffea* L. contains two subgenera: *Eucoffea* which comprises the caffeine-containing coffee-trees from Africa and *Mascarocoffea* in which caffeine-free coffee-trees from the Malagasy region are grouped (Chevalier 1938, 1942). The subgenus *Eucoffea*, originating from central and east African centers, is differentiated into *C. arabica* from Ethiopia, *C. eugenioides*

from Kenya and *C. canephora* as well as *C. liberica* from west Africa, especially the Ivory Coast. Some other species like *C. congensis*, *C. excelsa* and "nana" taxon come from the Central African Republic.

The species of the *Eucoffea* subgenus are wild or cultivated. The most important commercial species, *C. arabica* covering about 70% of the plantations, is the only tetraploid ( $2n=44$ ) known in the genus and is self fertile. The other members of the *Eucoffea* section are all diploids ( $2n=22$ ) and generally self incompatible. *C. canephora*, commonly called robusta and covering about 26% of the plantations, is the only diploid species of the genus whose history as cultivated plant can be compared with the history of *C. arabica*. It has developed world wide as a species for industrial plantation following the decline of the cultivation of *C. arabica* in lower altitudes of the old world after the advent of the coffee leaf rust *Hemileia vastatrix*.

One of the basic problems in the phylogeny of the coffee trees concerns the origin of the tetraploid *C. arabica*. This species exhibits the cytogenetical behaviour of allotetraploid plants (Carvalho 1952) suggesting that it has evolved from a cross between two wild diploid species. One nuclear genome could have originated from any member of the genus *Coffea*, the other genome could be sought among the wild species of the subgenus *Eucoffea* or in the neighbouring genera *Paracoffea* J. F. Ler. or *Psilanthus* Hook. f. (Charrier 1978). Conceivably, *C. eugenioides* could be the mother and *C. congensis* or *C. canephora* the father. This hypothesis is supported by allozymic variation in a wide range of wild populations and a preliminary comparison of the mt DNA of three *Eucoffea* species (Berthou and Trouslot 1977; Berthou et al. 1980).

The specific cleavage of chloroplast (cp) and mitochondrial (mt) DNAs by restriction endonucleases and agarose slab gel electrophoresis of the resulting DNA fragments have been used previously to probe the taxonomic relationships in the genera *Triticum* (Vedel et al. 1978, 1980), *Zea* (Timothy et al. 1979), *Nicotiana* (Rhodes et al. 1981; Kung et al. 1982), *Oenothera* (Gordon et al. 1982), *Pelargonium* (Metzlaff et al. 1981) and *Brassica* (Lebacqz and Vedel 1981). These studies have demonstrated a diversity among cytoplasmic genomes

12 DEC. 1983

O. R. S. T. O. M. Fonds Documentaire

N° : 4205 ex 1

Cote : B

B4205 ex 1

which was more pronounced as the taxonomic plant groups diverged. We present here a restriction enzyme analysis of cytoplasmic DNAs from nine coffee trees to explore phylogenetic relationships in the *Eucoffea* subgenus and consequently the origin of the tetraploid *C. arabica*.

## Material and methods

### Species examined

The geographical origins of the nine analysed taxons are listed in Table 1. — All species but one belong to the subgenus *Eucoffea*. The species *Paracoffea ebracteolata* belongs to the genus containing caffeine-free shrubs and has been previously analysed for its allozymes (Berthou and Trouslot 1977).

### Organelle isolation

Chloroplasts and mitochondria were isolated from flushes of wet-summer grown leaves collected on trees in plantations and clone orchards. Coffee tree leaves generally are waxy and stiff and furthermore contain tannins and polyphenols. Consequently, important modifications have been introduced both in the homogenization and in the purification process of the crude chloroplast pellets.

The homogenization of leaves was carried out at 4°C with a Waring blender, 3 × 10 s at maximum speed, in a buffer containing: 0.3 M mannitol, 0.05 M tris, 0.15 M ascorbic acid, 0.1 M mercaptoethanol, 3 × 10<sup>-3</sup> M EDTA, 0.1% BSA pH 8.0. The homogenate was filtered through a 20 µm nylon net and the filtrate centrifuged at 1,000 g, 12 min (IEC, rotor 269). The pellet consisted of the crude chloroplast fraction. The supernatant was spun at 20,000 g, 15 min (Sorvall, rotor SS 34); the pellet consisted of the crude mitochondrial fraction. DNase I treatment was omitted in the case of the crude chloroplast frac-

tions for two reasons: (i) these fractions are composed of broken organelles so that DNase treatment destroys all the cp DNA, (ii) the restriction enzyme analysis does not show cleaved fragments of chromatin from nuclear contaminants. The crude chloroplast pellets were gently dispersed in A buffer (Herrmann et al. 1975) and fractionated by centrifugation at 4°C in discontinuous sucrose gradients 30%, 60% steps in A buffer at 2,000 g, 20 min (IEC, rotor 269). Material banding at the interface was carefully removed, slowly diluted with 3 volumes of A buffer and pelleted at 1,200 g, 12 min. The crude mitochondrial pellets were dispersed in A buffer with a Potter homogenizer and the mitochondrial fractions purified successively by DNase-treatment (Herrmann et al. 1975) and centrifugation at 4°C in discontinuous sucrose gradients, 30%, 60% steps in A buffer at 60,000 g, 1 h (Beckman, rotor SW 27 II). Material banding at the interphase was collected as indicated above with chloroplasts.

### DNA restriction and agarose gel electrophoresis

Cp and mt DNAs were isolated from the corresponding purified fractions by using CsCl-ethidium bromide gradients as described previously (Vedel et al. 1980). Sal I was prepared by the method of Greene et al. (1978); Bgl I, Xho I and Hpa II were purchased from Boehringer, Mannheim; Kpn I was purchased from New England Biolabs. Cp and mt DNAs were digested with restriction endonucleases as described by the suppliers. The restriction fragments were separated by electrophoresis in 0.7% agarose vertical slab gels, 20 or 40 cm long. Gel staining and ultraviolet fluorescence photography have been described (Vedel et al. 1976; Quéfier and Vedel 1977). A mixture of DNA fragments generated from λ DNA by Hind III and from λ DNA by Hind III + Eco RI (Boehringer, Mannheim) was used as a molecular weight standard.

### Analytical ultracentrifugation

DNA fractions recovered from CsCl-ethidium bromide gradients were submitted to analytical ultracentrifugation (Beck-

**Table 1.** Coffee-tree populations or clones used in experiments (P = population, C = clone)

Species or taxon	Entry or source	Prospector, country and site of original source
<i>C. arabica</i>	P8 P36 <sub>2</sub>	Guillaumet and Hallé 1967, Ethiopia: Wush-Wush Kollo districts
<i>C. eugenioides</i>	P1 P2	Berthaud and Guillaumet 1980, Kenya: Kakamega, Nandi districts
<i>C. congensis</i>	P1 P2	Berthaud and Guillaumet 1978, Central African Republic: Oubangui river
<i>C. canephora</i>	C 197 C 315	Portères 1937, Ivory Coast: Man Préfecture
"nana" taxon	C 91	Berthaud and Guillaumet 1978, Central African Republic: Nana-Nambéré Préfecture
<i>C. excelsa</i>	C 301	Berthaud and Guillaumet 1978, Central African Republic: Sobou village
<i>C. liberica</i> var. <i>indeniensis</i>	C 70	Robinet 1954, Ivory Coast: Indénié region
<i>Paracoffea ebracteolata</i>	P1	Le Pierrès 1976, Ivory Coast: Divo
<i>C. arabusta</i>	<i>C. canephora</i> ♀ × <i>C. arabica</i>	Capot 1972, Ivory Coast: Bingerville

man, model E ultracentrifuge) after the dye had been removed by gel filtration. Equilibrium centrifugation of 2–5  $\mu$ g of DNA in neutral CsCl solutions ( $\rho = 1,700$  g/cm<sup>3</sup>) was performed at 44,000 rpm for 20 h at 20 °C. *Micrococcus lysodeikticus* DNA, density 1,731 g/cm<sup>3</sup> was included with the sample as a density standard. Buoyant densities and base composition were calculated according to the method of Schildkraut et al. (1962).

## Results

### Restriction analysis of coffee-tree cp DNAs

Figure 1 illustrates the restriction patterns of *C. arabica* cp DNA generated by five different enzymes, from the least to the most complicated patterns: Sal I (5 bands), Kpn I (8 bands), Bgl I (9 bands), Xho I (10 bands) and Hpa II (more than 25 bands). The largest uv absorbing band in each pattern represents nuclear DNA contamination. Unfortunately, the DNase treatment of the chloroplast pellets previously used to remove selectively the nuclear DNA cannot be applied in the case of coffee trees. Because of their peculiar structure, coffee-tree leaves lead even after mild homogenization to pellets of broken chloroplasts. Consequently, DNase treatment of such pellets cleaves chloroplast DNAs.

Hpa II restriction enzyme fragmentation gives the most useful pattern to compare coffee-tree cp DNAs because (1) this enzyme produces the largest number of cp DNA fragments among the enzymes used and therefore has the highest resolving power to uncover differences; (2) the low molecular weight of the Hpa II fragments prevents interference with the nuclear DNA contamination.

In the genus *Coffea*, the Hpa II restriction patterns for cp DNA fall into three sub-sets. They are from the species: (i) *C. arabica* and *C. eugenioides* (Fig. 2, lanes a and b); (ii) *C. canephora*, *C. congensis* and "nana" taxon (Fig. 2, lanes c, d and e); (iii) *C. liberica* (Fig. 2, lane f). There is no difference between *C. arabica* and *C. eugenioides* cp DNAs nor between *C. canephora* and *C. congensis*. Among the related species of *C. arabica* in the *Erythrocoffea* A. Chev. sub-section and at the junction of the *Abyssinicae-Robustae* series J. Lebrun, this analysis provides evidence for two phylogenetic subdivisions.

The three subsets are distinguished by comparison of the nineteen heaviest Hpa II bands (Table 2). Subset I is characterized by cp DNAs containing two bands of  $2.3$  and  $0.83 \times 10^6$  d in size (arrows in Fig. 2 and Table 2) which are absent in the other two subsets. Subset II contains one unique fragment,  $1.53 \times 10^6$  d in molecular weight (arrow in Fig. 2c), and subset III, two unique ones  $1.77$  and  $1.48 \times 10^6$  d in size (Fig. 2f). Furthermore, the cp DNAs of the subset I lack the  $1.15$  and  $0.80 \times 10^6$  d bands which are present in the cp DNAs of the subsets II and III (asterisks in Fig. 2c and Table 2).

### Sal I restriction analysis of coffee-tree mt DNAs

The situation encountered with coffee-tree mt DNAs is quite different from that described above with cp DNAs because (1) the nuclear DNA contamination can be removed easily by DNase treatment of the mitochondrial pellets without loss of mt DNA, (2) restriction enzyme cleavage of mt DNAs from coffee-trees leads to very numerous fragments. Among the restriction patterns ob-

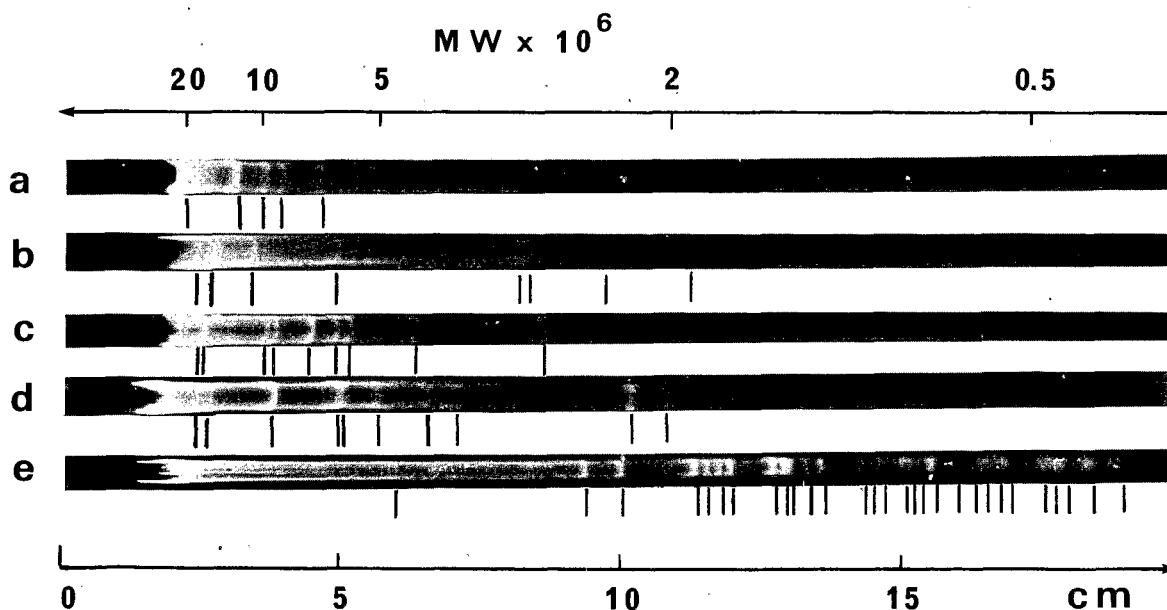
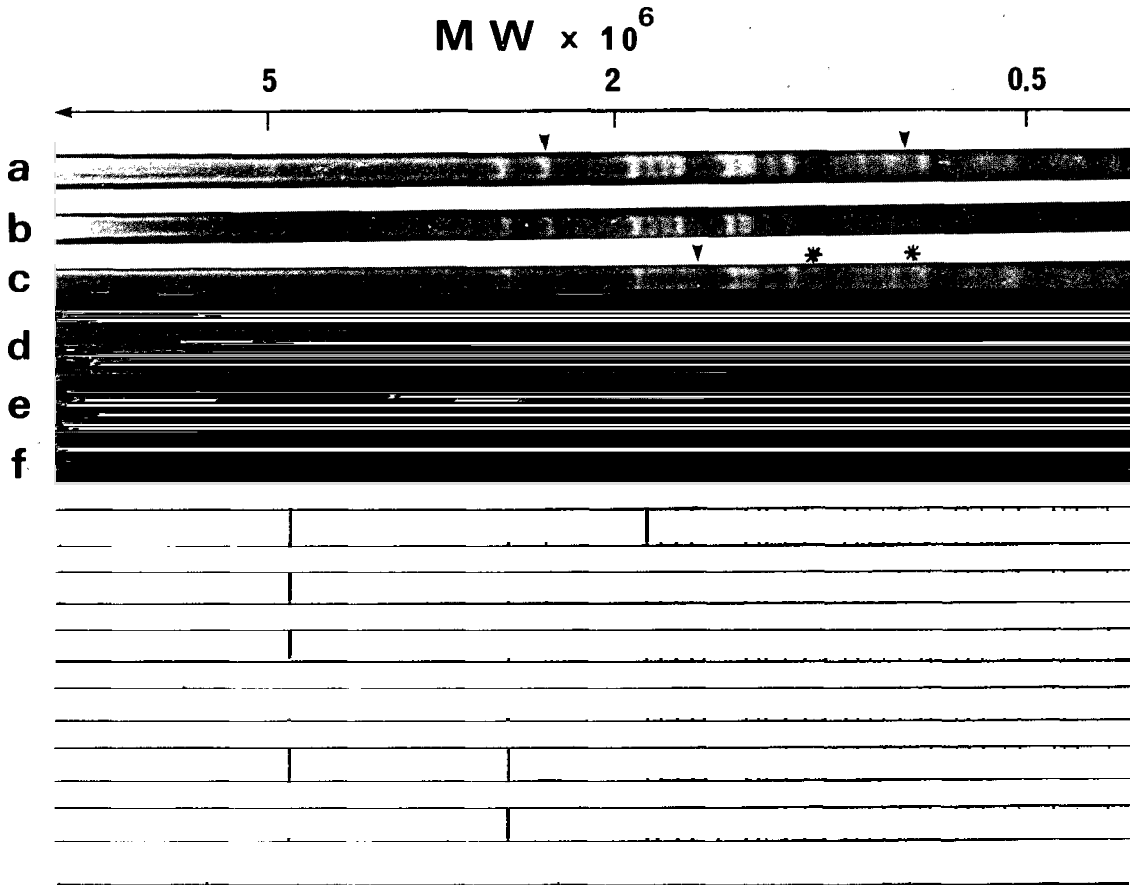


Fig. 1. Agarose slab gel electrophoresis of *Coffea arabica* cp DNA digested with several enzymes: a Sal I; b Kpn I; c Bgl I; d Xho I; e Hpa II



tained, the most useful ones are the Sal I patterns because they contain the smallest number of bands.

The mt DNA patterns present about twelve times more restriction fragments than the corresponding cp DNA by Sal I (Fig. 3, lane j). The purity of the mt DNA from *C. congensis* has been further checked by analytical CsCl gradients. The presence of a single and sharp band at  $1.706 \text{ g/cm}^3$  (Fig. 4) indicates that nuclear DNA is completely removed by the DNase treatment of the mitochondrial pellets and that cp DNA is not responsible for the complexity of the mt DNA pattern.

Comparison of the Sal I patterns of the mt DNAs isolated from eight coffee-tree taxons revealed five sub-sets. Overall similarities are observed among *C. arabica*, *C. eugenioides* and *C. congensis* as among *C. canephora* and "nana" taxon (Fig. 3, lane b-f). In contrast, *C. excelsa*, *C. liberica* and *P. ebracteolata* are characterized by uniquely differing mt DNA patterns (Fig. 3, g, h, i). The complexity of the patterns precludes accurate de-

termination of restriction fragment homologies between mt DNAs of the five sub-sets. The molecular weight of the mt DNAs of *C. arabica*, *C. eugenioides*, *C. congensis* and *C. canephora* estimated by summing the molecular weight of the fifty heaviest bands (above  $1.2 \times 10^6 \text{ d}$ ) results in a value greater than  $360 \times 10^6 \text{ d}$  in all restriction patterns. This value does not take into account band multiplicity. Thus estimated sizes of the coffee-tree mt DNAs appear greater than the sizes of other analysed higher plant mt DNAs, except for some *Cucurbitaceae* (Ward et al. 1981).

#### *Maternal inheritance of C. arabusta cytoplasmic DNAs*



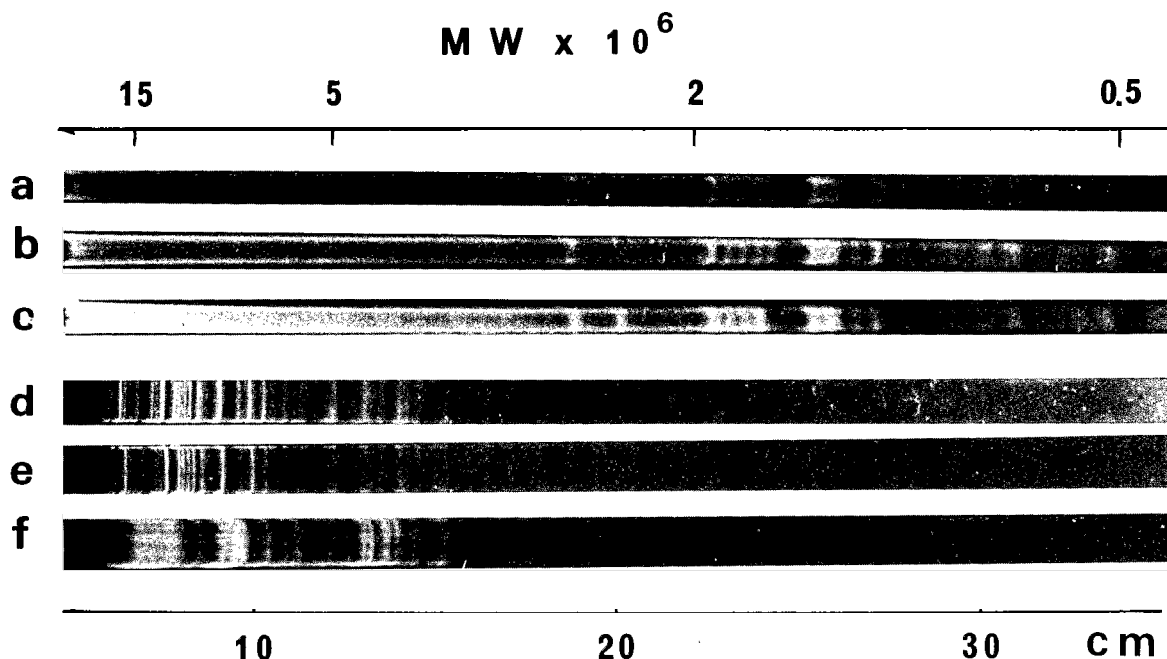


Fig. 5. Agarose slab gel electrophoresis of Hpa II digests of cp DNA from: a *Coffea canephora*; b *C. arabusta*; c *C. arabica* and of Sal I digests of mt DNAs from: d *C. canephora*; e *C. arabusta*; f *C. arabica*

Table 2. Molecular weights ( $\times 10^6$  d) and distribution of the nineteen heaviest Hpa II restriction fragments of coffee-tree cp DNAs

Subset I	Subset II	Subset III
<i>C. arabica</i>	<i>C. congensis</i>	<i>C. liberica</i>
<i>C. eugenioides</i>	<i>C. canephora</i> "nana" taxon	
5.25	5.25	5.25
2.69	2.69	2.69
2.30		
1.81	1.81	1.81
		→1.77
1.75	1.75	
1.65	1.65	1.65
1.58	1.58	1.58
	→1.53	
		→1.48
1.40	1.40	1.40
1.35	1.35	1.35
1.32	1.32	1.32
1.27	1.27	1.27
1.23	1.23	1.23
	* 1.15	* 1.15
1.05	1.05	1.05
1.00	1.00	1.00
0.95	0.95	0.95
0.90	0.90	0.90
0.86	0.86	0.86
→0.83		
	* 0.80	* 0.80
0.77	0.77	0.77

lated to *C. arabica* (Lebrun 1941). *C. canephora* and *C. congensis* may have a common origin corresponding to the *robustoides* A. E. Haarer (1957) series as judged by the cp DNA pattern; but these species are slightly different with regard to the mt DNA patterns. This suggests the divergence of this phylogenetic cluster before the separation of *C. arabica* and *C. eugenioides*. *C. arabica* and *C. eugenioides* are included in a series different from *Abys-sinicae* J. Lebrun with the exclusion of *C. congensis*. The inclusion of the latter species would be correct as far as the mt DNA pattern is concerned (Table 3).

(iii) Species can be distinguished by the organelle DNA. *C. excelsa* A. Chev. from the Central African Republic and *C. liberica* Bull ex Hiern var. *indeniensis* Sibert from the Ivory Coast are distinct. In J. Hutchinson and J. M. Dalziel Flora (Hutchinson and Dalziel 1963), R. W. J. Keay gathers them under the binomial *C. liberica* Bull ex Hiern. The mt DNA patterns do not agree with this view for they show a clear-cut difference between these species.

Finally, the "nana" taxon from the Central African Republic is identical to *C. canephora* from the Ivory Coast, whereas we observe mt DNA divergence between *C. canephora* and *C. congensis*. We can conclude that the "nana" taxon belongs to the species *C. canephora*.

From an ecological point of view, the two phylogenetic clusters gather the orophytic coffee-trees on the

**Table 3.** Comparison of systematic series or subgenera and observed groups following cp and mt DNA patterns

A. Chevalier 1947	J. Lebrun 1941	A. E. Haarer 1957	DNA cp	DNA mt
<i>Erythrocoffea</i> <i>C. arabica</i> <i>C. eugenioides</i> <i>C. congensis</i>	} <i>Abyssinicae</i>	}	} Group 1	→ Group A
		} A. E. Haarer's series	→ Group 2	→ Group A'
<i>C. canephora</i>	→ <i>Robustae</i>			

one hand, i.e. *C. arabica* from high valleys of the Ethiopian plateau and *C. eugenioides* from the lowland forest of the Lake Victoria Belt (Guillaumet and Hallé 1967; Berthaud et al. 1980) and the lower altitude coffee trees on the other hand, i.e. *C. congensis*, a riparian species and *C. canephora*, a very plastic one (Chevalier 1947; Berthaud and Guillaumet 1978). The former phylogenetic cluster is not differentiated, but the latter is differentiated Eastwards and Westwards between *C. congensis* and *C. canephora* on the basis of mt DNA structure. But we can notice that another kind of differentiation Northwards and Southwards must be analysed considering the species *C. zanguebariae*, *C. nufindiensis* and *C. racemosa*.

These results support the hypothesis of the formation of *C. arabica* once the combination of the two *Eucocoffea* main phylogenetic clusters had occurred. *C. arabica* diverged from an ancestor similar to *C. eugenioides* and this was preceded by the divergence of *C. canephora* and *C. congensis*. This conclusion agrees with previous isozyme studies. So when plant breeders want to extend the selection basis of *C. arabica* using more resistant diploid coffee-trees from lower altitude (coffee leaf rust, drought), they should be aware of the cytoplasmic DNA differentiation. As far as this differentiation is concerned, the diploid species *C. eugenioides* and *C. congensis* can play differential roles as female parents in hybrids.

**Acknowledgements.** We thank D. Le Pierres, F. Anthony, E. Couturon and Dr. A. Charrier for sending samples of fresh coffee-trees leaves from Ivory Coast. We also acknowledge Mrs. M. O. Mosse for analytical ultracentrifugation of mitochondrial DNA samples.

## References

- Berthaud J, Guillaumet JL (1978) Les caféiers sauvages en Centrafrique. Résultats d'une mission de prospection (Janvier-Février 1975). *Café Cacao The* 22:171-186
- Berthaud J, Guillaumet JL, Le Pierres D, Lourd M (1980) Les caféiers sauvages du Kenya: prospection et mise en culture. *Café Cacao The* 24:101-112
- Berthou F, Trouslot P (1977) L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une méthode d'électrophorèse en série; premiers résultats. 8ème Colloque Scientifique International sur le café, Abidjan, ASIC, Paris 1979, pp 373-384
- Berthou F, Trouslot P, Hamon S, Vedel F, Quetier F (1980) Analyse en électrophorèse du polymorphisme biochimique des caféiers: variation enzymatique dans dix-huit populations sauvages; variation de l'ADN mitochondrial dans les espèces *C. canephora*, *C. eugenioides* et *C. arabica*. *Café Cacao The* 24:313-326
- Carvalho A (1952) Taxonomia de *Coffea arabica* L. 6. Caracteres morfológicos dos haploides. *Bragantia* 12:201-212
- Charrier A (1978) La structure génétique des caféiers spontanés de la région malgache *Mascarocoffea*. Leurs relations avec les caféiers d'origine africaine (*Eucocoffea*). *Mem ORSTOM* 87:223
- Chevalier A (1938) Essai d'un groupement systématique des caféiers sauvages de Madagascar et des Iles Mascareignes. *Rev Bot Appl Agric Trop* 18:825-843
- Chevalier A (1942) Les caféiers du globe. 2. Iconographie des caféiers sauvages et cultivés et des Rubiacées prises pour des caféiers. *Encycl Biol P. Lechevalier*, Paris
- Chevalier A (1947) Les caféiers du globe. Systématique des caféiers et des faux-caféiers, *Encycl Biol* 28. P. Lechevalier, Paris, pp 82-263
- Guillaumet JL, Halle F (1967) Etude de la variabilité du *Coffea arabica* dans son aire d'origine. *Rapport ORSTOM*
- Greene PJ, Heynecker HL, Bolivar F, Rodriguez RL, Betlach MC, Covarrubias AA, Backman K, Russel DJ, Tait R, Boyer HW (1978) A general method for the purification of restriction enzymes. *Nucleic Acids Res* 5:2373-2380
- Gordon KHJ, Crouse EJ, Bohnert HJ, Herrmann RG (1982) Mapping of differences in chloroplast DNA of five wild-type plastomes in *Oenothera* subsection *Euoenothera*. *Theor Appl Genet* 61:373-384
- Haarer AE (1957) Economic species varieties and forms of coffee. *World Crops*. 9:371-373
- Herrmann RG, Bohnert HJ, Kowallik KV, Schmitt JM (1975) Size, conformation and purity of chloroplast DNA from some higher plants. *Biochim Biophys Acta* 378:305-317
- Hutchinson J, Dalziel JM (1963) *Flora of West Tropical Africa* (2nd edn by Heppen FN) 2:152-157
- Kung SD, Zhu YS, Shen GF (1982) *Nicotiana* chloroplast genome. 3. Chloroplast DNA evolution. *Theor Appl Genet* 61:73-79
- Lebacqz P, Vedel F (1981) Sal I restriction enzyme analysis of chloroplast and mitochondrial DNAs in the genus *Brassica*. *Plant Sci Lett* 23:1-9

- Lebrun J (1941) Recherches morphologiques et systématiques sur les caféiers du Congo. Publ Inst Natl Etude Agron Congo, Ser Tech
- Metzlaff M, Börner T, Hagemann R (1981) Variations of Chloroplast DNAs in the Genus *Pelargonium* and their Biparental inheritance. Theor Appl Genet 60:37-41
- Quétier F, Vedel F (1977) Heterogeneous population of mitochondrial DNA molecules in higher plants. Nature 268:365-368
- Rhodes PR, Zhu YS, Kung SD (1981) *Nicotiana* chloroplast genome. I. Chloroplast DNA diversity. Mol Gen Genet 182:106-111
- Schildkraut CL, Marmur J, Doty P (1962) Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl. J Mol Biol 4:430-443
- Timothy DH, Levings CS, Pring DR, Conde MF, Kermicle JL (1979) Organelle DNA variation and systematic relationships in the genus *Zea*: Teosinte. Proc Natl Acad Sci USA 76:4220-4224
- Vedel F, Quétier F, Bayen M (1976) Specific cleavage of chloroplast DNA from higher plants by Eco RI restriction nuclease. Nature 263:440-442
- Vedel F, Quétier F, Dosba F, Doussinault G (1978) Study of wheat phylogeny by Eco RI analysis of chloroplastic and mitochondrial DNAs. Plant Sci Lett 13:97-102
- Vedel F, Lebacqz P, Quétier F (1980) Cytoplasmic DNA variation and relationships in cereal genomes. Theor Appl Genet 58:219-224
- Ward BL, Anderson RS, Bendich AJ (1981) The mitochondrial genome is large and variable in a family of plants (*Cucurbitaceae*). Cell 25:793-803