Theor Appl Genet (1998) 97: 1204–1209

S. N. Raina · Y. Mukai · M. Yamamoto

In situ hybridization identifies the diploid progenitor species of *Coffea arabica* (Rubiaceae)

Received: 2 February 1998 / Accepted: 12 May 1998

Abstract The most important commercial coffee species, *Coffea arabica*, which is cultivated in about 70% of the plantations world-wide, is the only tetraploid ($2n = 4x = 44$) species known in the genus. Genomic in situ hybridization (GISH) and fluorescent in situ hybridization (FISH) were used to study the genome organization and evolution of this species. Labelled total genomic DNA from diploid species (*C. eugenioides*, *C. congensis*, *C. canephora*, *C. liberica*) closely related to *C. arabica* was separately used as a probe in combination with or without blocking DNA to the chromosome spreads of *C. arabica*. GISH discriminated between chromosomes of *C. arabica* only in the presence of an excess of unlabelled block DNA from the species not used as a probe. Among the range of different species combinations used, DNA from *C. eugenioides* strongly and preferentially labelled 22 chromosomes of the tetraploid *C. arabica*, while the remaining 22 chromosomes were labelled with *C. congensis* DNA. The similarity of observations between *C. arabica* and the two diploid species using two ribosomal genes with FISH with respect to metaphase chromosomes provided additional support to the GISH results. These results confirm the allopolyploid nature of *C. arabica* and show that *C. congensis* and *C. eugenioides* are the diploid progenitors of *C. arabica*.

Key words *Coffea arabica* · In situ hybridization · Genomic probe · Ribosomal DNA probes · Allopolyploidy · Wild progenitors

Communicated by B. S. Gill

S. N. Raina[1] (✉) · Y. Mukai
Laboratory of Plant Molecular Genetics,
Division of Natural Science, Osaka
Kyoiku University, 4-698-1, Asahigaoka,
Kashiwara, Osaka 582, Japan

M. Yamamoto
Department of Life Sciences, Kansai Women's College,
3-11-1, Asahigaoka, Kashiwara, Osaka 582, Japan

Present address:
[1] Laboratory of Cellular and Molecular Cytogenetics,
Department of Botany, University of Delhi,
Delhi 110 007, India

Introduction

Coffee plays a major role in the economy of many African, American and Asian countries. It is now grown commercially in more than 50 countries with a turnover of about 18 billion US dollars (Orozco-Castilho et al. 1994). There are many species (approx. 50) of coffee, but only 2 are important commercially, *Coffea canephora* and *C. arabica*; these are responsible for the total world production. The better quality (low content of caffeine and a fine aroma) of *C. arabica* or Arabica Coffee makes it by far the most important species commercially. It represents 73% of the world production and almost all of the production in Latin America. With the exception of *C. arabica*, all of the coffee species are diploids ($2n = 2x = 22$). *C. arabica* is a natural, self-fertile tetraploid ($2n = 4x = 44$), while diploids are generally self-sterile, allogamous species (Charrier and Berthaud 1985; Dublin et al. 1991).

C. arabica exhibits the cytogenetical behaviour of allotetraploid plants (Carvalho 1952; Charrier and Berthaud 1985), suggesting that it has evolved from a cross between two wild diploid species. One question of general interest, therefore, is the identity of the progeniter species that hybridized to form cultivated *C. arabica*. Despite its economic and agricultural importance, little genetic research has been devoted to identify the sources of its component genomes that would ensure the most efficient use of wild species' germplasm in the improvement of the crop. This is all the more

important in the context of its susceptibility to various pests and diseases and very low level of genetic variability (Wrigley 1995; Dublin et al. 1991; Paillard et al. 1996). The diploid species closely related to *C. arabica*, on the other hand, have wide genetic diversity and possess agronomically useful characters of importance (Dublin et al. 1991; Wrigley 1995).

With the advance of in situ hybridization, the possibilities of elucidating parental genome contributions in allopolyploids and molecular karyotypes, investigating nuclear organization and physically mapping DNA sequences to both metaphase and meiotic chromosomes and interphase chromatin have considerably increased. The use of total genomic DNA as a probe in hybridization experiments to chromosomal DNA in situ (genomic in situ hybridization, GISH) (Schwarzacher et al. 1989) is particularly useful in characterizing the genomes and chromosomes of hybrid plants, allopolyploid species and recombinant breeding lines (Anamthawat-Jonsson et al. 1990; Bailey et al. 1993; Bennett et al. 1992; Chen and Armstrong 1994; Kenton et al. 1993; Mukai and Gill 1991; Parokonny et al. 1994; Schwarzacher et al. 1992; Thomas et al. 1994; Jiang et al. 1994; Heslop-Harrison et al. 1990). Here we describe evidence in support of the allopolyploid origin of *C. arabica* and establish for the first time the identity of the ancestor genome donor species in *C. arabica* using GISH and chromosome markers.

Materials and methods

Plant material

Table 1 shows the species used in the experiments. Besides *C. arabica*, all the diploid species listed in subsection Erythrocoffea as well as *C. liberica* (all considered to be closely related to *C. arabica*) from subsection Pachycoffea were included in this study. The seeds were sown in pots in the greenhouse and started germinating only approximately 60 days after sowing.

Table 1 List of coffee species used in the present study

Species	Section (subsection)	2n	Status
C. congensis	Eucoffea (Erythrocoffea)	22	Wild
C. eugenioides	Eucoffea (Erythrocoffea)	22	Wild
C. canephora	Eucoffea (Erythrocoffea)	22	Cultivated
C. arabica	Eucoffea (Erythrocoffea)	44	Cultivated
C. liberica	Eucoffea (Pachycoffea)	22	Wild

Chromosome preparations for in situ hybridization

Young growing root tips were immersed in distilled water and cold-treated at 4°C for 20 h. They were then fixed in freshly prepared 1:3 glacial acetic acid: absolute alcohol for at least 24 h before staining in 1% aceto-carmine for 15 min. Dissected meristems were squashed in 45% acetic acid. The slides were quickly placed on a slab of dry ice for 10 min before removal of the coverslips with a razor blade. The slides were immersed in a staining jar containing 45% acetic acid for 15 min, air-dried, and stored in a dessicator at − 20°C until used.

DNA isolation and probe preparation

Total genomic DNA was isolated from seedling leaves according to the procedure of Saghai Maroof et al. (1984). DNA concentration was measured by a Gilford spectrophotometer at a wavelength of 260 nm.

The genomic DNA (1 μg/reaction) from each diploid species was labelled with biotin-16-dUTP using the BioNick Labelling System (Boehringer-Mannheim) as instructed by the supplier.

The ribosomal gene-specific probes used in this study were pTa 71 (Gerlach and Bedrook 1979) and pTa 794 (Gerlach and Dyer 1980). pTa 71 contained a 9.05-kb *Eco*RI fragment of a full-length nuclear rDNA repeat unit (18S-5.8S-26S genes and spacers) of wheat; pTa 794 consisted of 410-bp *Bam*HI fragment of 5S rDNA isolated from wheat. pTa 71 was labelled with fluorescein isothiocyanate (FITC)-dUTP conjugate, and pTa794 was labelled with digoxigenin-11-dUTP by nick translation.

In situ hybridization and detection

Chromosomal DNA on the slide was denatured at 69°C for 2 min in 50 ml of 70% formamide, 2 × SSC. The slides were then rapidly dehydrated in an ethanol series (70%, 95%, 100%, 5 min each) at − 20°C and air-dried. The probe mix (100 μl total volume for ten slides) containing 50% (v/v) de-ionized formamide, 10% (w/v) dextran sulphate, 2 × SSC, 1 μg of biotin-labelled total genomic DNA, 25 μg of unlabelled autoclaved genomic DNA for blocking (optional) and 50 μg of sonicated salmon sperm DNA was denatured at 100°C for 10 min and quickly cooled on ice for 10 min. Ten microlitres of denatured probe mix was applied onto the denatured slide preparation and covered with a coverslip. The slides were then placed in a humid chamber and incubated at 37°C overnight.

After hybridization overnight, the slides were immersed in 2 × SSC at room temperature for 5 min to loosen the coverslips. The slides were then washed by immersing in 50% (v/v) de-ionized formamide in 2 × SSC at 37°C for 15 min, 2 × SSC at room temperature for 15 min, 1 × SSC at room temperature for 15 min and 4 × SSC at room temperature for 5 min. Detection of the biotinylated probe was achieved using fluorescein-conjugated streptavidin (Boehringer Mannheim). Slides were incubated in 10 μg/ml fluorescein-conjugated streptavidin in 4 × SSC – 1% BSA for 1 h at 37°C. After incubation, the slides were washed in 4 × SSC for 10 min, 0.1% Triton X-100 in 4 × SSC for 10 min, 4 × SSC for 10 min, and 2 × SSC for 5 min, all at room temperature. Twenty-five microlitres of fluorescence anti-fade solution (Vectashield, Vector Laboratories) containing non-specific fluorochrome propidium iodide (1 μg/ml) was applied to each slide. The slides were covered with a coverslip and placed in dark for 1 h at room temperature. The in situ hybridization signal and propidium iodide were excited with blue light (450–490 nm) and detected by their yellow and red fluorescence, respectively.

For in situ hybridization using pTa 71 and pTa 794 as probes, all conditions were same as those described above except that the two

probes were added to the hybridization mix at concentrations of 1.0 and 0.5 μg/100 μl, respectively. For detection, the slides were incubated in 10 μg/ml anti-dig rhodamine in 4 × SSC-1% BSA, and chromosomes were counterstained with differentiating fluorochrome DAPI (4′-6-diamidino-2-phenylindole). The in situ hybridization signals and DAPI fluorescence were visualized with appropriate filters.

Results and discussion

Initially, no blocking DNA was used for the first probing. The results enabled us not only to develop a strategy for ordering GISH analysis but also to gain initial information about the physical distribution of sequences which are common or different between the species being probed (*C. arabica*) and the species used to supply the probe DNA. When the *C. eugenioides* or *C. congensis* probe was applied on the chromosome preparations of *C. arabica*, there was considerable cross hybridization of these species probes to the *C. arabica* chromosomes. However, even without blocking, it was clear that the DNA extracted from either species hybridized more strongly to the alternative subsets (22) of chromosomes of *C. arabica*.

In subsequent experiments, total genomic DNA from one species was used as the labelled probe, while unlabelled DNA from the other species was applied at a much higher concentration as a block, so that the blocking DNA would hybridize to sequences in common between the block and the labelled probe and mainly genome-specific sequences would remain as sites for probe hybridization (Anamthawat et al. 1990). This enables closely related genomes to be distinguished by GISH (Bennett 1995).

Probing the *C. arabica* chromosome spreads with biotin-labelled total genomic DNA derived from *C. congensis* together with an excess of unlabelled genomic DNA from *C. eugenioides* resulted in strong probe hybridization in the 22 chromosomes of *C. arabica* (Fig. 1a); the remaining 22 chromosomes were almost totally unlabelled. In a reciprocal GISH using labelled *C. eugenioides* as a probe with an excess of unlabelled *C. congensis* DNA block, the reverse was obtained (Fig. 1b). Labelled chromatin fluoresced yellow, while unlabelled chromatin fluoresced red with non-specific counterstain propidium iodide. The label in both cases was almost uniform, while distal regions showed little or no label.

Genomic DNA contains all types of repetitive and single-copy DNA sequences, and the proportion of the former increases with genome size (Raina and Narayan 1984; Bennett 1995); for example, from 40% in *Arabidopsis thaliana* (2C = approx. 0.1 pg) to more than 95% in *Allium cepa* (2C = 33.4 pg). It is very unlikely that many of the single-copy sequences find their homologous partners under the in situ hybridization conditions used in the present experiments. Hence, the signal seen by us most likely arose from hybridization between widely dispersed members of families of highly and middle repetitive sequences and that the genomic hybridization method largely examined the organization of these sequences (Bennett 1995). The 44 chromosomes of *C. arabica* are very small in size. The average genome size of *C. arabica* is appreciably smaller (2C = 2.5 pg, Bennett and Leitch 1995) than those of many of the taxa analyzed by GISH so far, and so the low abundance of dispersed repetitive DNA sequences within *C. arabica* chromosomes at the distal regions is likely to be responsible for the few or no hybridization signals, as is the case for poor genome size species *Brassica* and *Arabidopsis* (Kamm et al. 1995; Heslop-Harrison and Schwarzacher 1996).

The other two species which are considered to be closely related to *C. arabica* are *C. canephora* and *C. liberica*. Disease resistance genes have been transferred into the *C. arabica* gene pool via natural and artificial hybridization involving these two species (Dublin et al. 1991; Wrigley 1995). We believed it of interest to find genome relationships between *C. arabica* and the two diploid species by GISH. Labelled total genomic DNA from *C. liberica* was used as a probe with an excess of unlabelled genomic DNA from *C. eugenioides* for in situ hybridization to the chromosome spreads of *C. arabica*. No sites of hybridization were detected on the chromosomes. This lack of detectable hybridization of the *C. liberica* genomic probe indicated that these chromosomes almost certainly derive from different taxa. Other studies have also shown that *C. arabica* is genetically distinct in its nuclear and organellar genomes from *C. liberica* (Berthou et al. 1980, 1983). In the presence of an excess of unlabelled genomic block DNA from *C. eugenioides* to the hybridization mixture, the labelled DNA from *C. canephora* hybridized to a subset of *C. arabica* chromosomes. The hybridization was, however, weak and uneven compared to the combination of *C. congensis* and *C. eugenioides*.

After several different combinations of probe and target DNAs were used among the coffee species, in situ hybridization using labelled total genomic DNA from *C. eugenioides* together with an excess of unlabelled DNA from *C. congensis*, and vice versa, showed that 22 chromosomes of *C. arabica* and *C. eugenioides* and the remaining 22 chromosomes of the former and *C. congensis* share a common origin. Thus, the present study confirms the allopolyploid origin of the naturally occurring *C. arabica* as previously suggested (Carvalho 1952; Charrier and Berthaud 1985). Based on restriction fragment length polymorphism (RFLP) analysis of chloroplast and mitochondrial DNAs and allozyme data, Berthou and Trouslot (1977) and Berthou et al. (1980, 1983) suggested that *C. eugenioides* could be the maternal and *C. congensis* or *C. canephora* the paternal progeniter of *C. arabica*. In the present study uniformly intense signals on 22 chromosomes of *C. arabica* probed with *C. congensis* in comparison to weak and

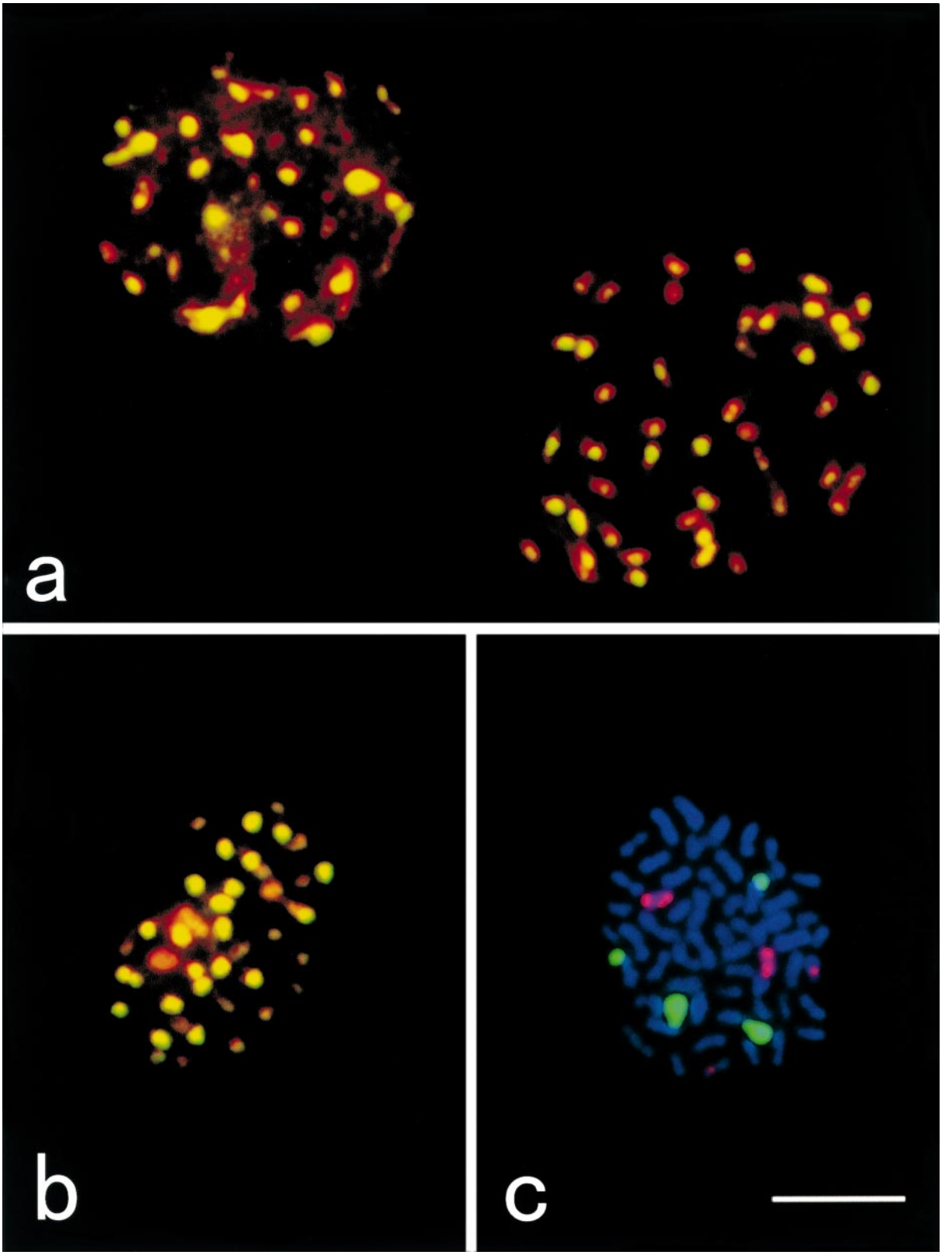

Fig. 1a, b Root-tip interphase (left) and mitotic chromosomes (right) of *Coffea arabica* following genomic in situ hybridization using biotinylated total genomic DNA from *C. congensis* as probe and unlabelled DNA from *C. eugenioides* as block (**a**), and biotinylated total genomic DNA from *C. eugenioides* as probe and unlabelled DNA from *C. congensis* as block (**b**). *Greenish-yellow* fluorescence indicates hybridization to the probe. Unlabelled chromatin fluoresces *reddish-orange* with propidium iodide. **c** Fluorescence in situ hybridization of *Coffea arabica* with the 18S-, 5.8S-, and 26S rDNA (*green* fluorescence) and 5S rDNA (*red* fluorescence) probes. Chromosomes were counterstained by DAPI

fragmented signals obtained with *C. canephora* leaves no doubt that *C. congensis* is the other progenitor species of *C. arabica*. The present results are in agreement with the conclusions reached by Berthou et al. (1980, 1983) about the close relationship between *C. canephora* and *C. congensis*.

Fluorescent in situ hybridization (FISH) using specific probes is also an effective method in supplementing information on the genome origins of polyploids (Fukui et al. 1994; Kamm et al. 1995; Parokonny and Kenton 1995; Linde-Laursen et al. 1992). In the present study, using pTa 71 and pTa 794 (labelled with FITC and digoxigenin, respectively) probes, bright probe hybridization sites were visible on chromosome spreads of *C. arabica*, *C. congensis* and *C. eugenioides*. Each of the three species have one pair of chromosomes with a secondary constriction (present study, Bouharmont 1959).

C. arabica had four pTa 71 hybridization sites (Fig. 1c). The largest two of these were associated with a secondary constriction (visualized by DAPI staining) in the 2 chromosomes. The remaining two were distinguished by having no secondary constriction and by showing medium-strength hybridization signals. The chromosome complement of *C. arabica* displayed six signals for pTa 794: four strong and two weak signals (Fig. 1c). Interestingly, each of the two large-sized signals were located at the distal ends of the same chromosome.

In situ hybridization using the pTa 71 probe identified 2 strong signals associated with as many nucleolar constrictions in the 2 chromosomes in both *C. congensis* and *C. eugenioides*. *C. eugenioides* like *C. arabica* was characterized by the presence of four large pTa 794 sites located at the distal ends of each of the 2 chromosomes. Localization of 5S loci to chromosome spreads revealed two weak hybridization signals on 2 chromosomes in *C. congensis*.

The similarity of observations between *C. arabica* and the two diploid species, especially the characteristic 5S rDNA loci located at distal ends of chromosomes in both *C. arabica* and *C. eugenioides*, support the view that *C. arabica* is a hybrid of *C. eugenioides* and *C. congensis*. Due to the small size of the median/submedian chromosomes and the small differences in total length and lengths of the long arms and short arms, respectively, the hybridization sites (except for nucleolar chromosomes) could not be assigned to the individual chromosomes and genomes.

Acknowledgements The first author (S.N.R.) is grateful, to the Indian National Science Academy and Japan Society for the Promotion of Science for financial support. Thanks are also due to Drs. Sreenath and Misra, Coffee Board, Mysore, India for the supply of seed samples, and to T. Yamada and Dr. M. K. Kaul for generous help and encouragement. This work was supported, in part, by a Grant-in-aid for Scientific Research (B) (No. 09490024) from the Ministry of Education, Science, Sports and Culture, Japan, and the Department of Biotechnology, Ministry of Science and technology, India.

References

Anamthawat-Jonsson K, Schwarzacher T, Leitch AR, Bennett MD, Heslop-Harrison JS (1990) Discrimination between closely related *Triticeae* species using genomic DNA as a probe. Theor Appl Genet 79: 721–728

Bailey JP, Bennett ST, Bennett MD, Stace CA (1993) Genomic in situ hybridization identifies parental chromosomes in the wild grass hybrid × *Festulpia hubbardii*. Heredity 71: 413–420

Bennett MD (1995) The development and use of genomic in situ hybridization (GISH) as a new tool in plant biosystematics. In: Brandham PE, Bennett MD (eds) Kew Chromosome Conf IV. Royal Botanic Gardens, Richmond, Surrey, pp 167–183

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. Ann Bot 76: 113–176

Bennett ST, Kenton AY, Bennett MD (1992) Genomic in situ hybridization reveals the allopolyploid nature of *Milium montianum* (Gramineae). Chromosoma 101: 420–424

Berthou F, Trouslot P (1977) L'analyse du polymorphisme enzymatique dans le genre *Coffea*: adaptation d'une méthode d'electrophorése en série; premiers résultats. In: Assoc Scient Intl Coffee (ASIC) (Eds) 8éme Colloque Sci Int Café. ASIC, Paris, pp 373–384

Berthou F, Trouslot P, Hamon S, Vedel F, Quetier F (1980) Analyse en électrophorése du polymorphisme biochimique des caféiers: variation enzymatique dans dix-huit populations sauvages; variation de l'ADN mitochondrial dans les espéces *C. canephora*, *C. eugenioides* et *C. arabica*. Cafe Cacao The 24: 313–326

Berthou F, Mathieu C, Vedel F (1983) Chloroplast and mitochondrial DNA variation as indicator of phylogenetic relationships in the genus *Coffea* L. Theor Appl Genet 65: 77–84

Bouharmont J (1959) Recherches sur les affinités chromosomiques dans le genre *Coffea*. Publ INEAC Ser Sci 77: 1–94

Carvalho A (1952) Taxonomia de *Coffea arabica* L. 6. Caracteres morfologicos dos haploides. Bragantia 12: 201–212

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 Held Publ, Beckenham, UK, pp 13–47

Chen Q, Armstrong K (1994) Genomic in situ hybridization in *Avena sativa*. Genome 37: 607–612

Dublin P, Enjalric F, Lardet L, Carron MP, Trolinder N, Pannetier C (1991) Estate crops. In: Debergh PC, Zimmerman RH (eds) Micropropagation: technology and practice. Kluwer Academic Publ, Dordrecht, The Netherlands, pp 337–361

Fukui K, Ohmido N, Khush GS (1994) Variability in rDNA loci in the genus *Oryza* detected through fluorescence in situ hybridization. Theor Appl Genet 87: 893–899

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7: 1869–1885

Gerlach WL, Dyer TA (1980) Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. Nucleic Acids Res 8: 4851–4865

Heslop- Harrison JS, Schwarzacher T (1996) Genomic southern and in situ hybridization for plant genome analysis. In: Jauhar PP (ed) Methods of genome analysis in plants, CRC Press, New York, pp 163–179

Heslop-Harrison JS, Leitch AR, Schwarzacher T, Anamthawat-Jonsson K (1990) Detection and characterization of 1B/1R translocations in hexaploid wheat. Heredity 65: 385–392

Jiang J, Morris KLD, Gill BS (1994) Introgression of *Elymus trachycaulus* chromatin into common wheat. Chromosome Res 2: 3–13

Kamm A, Galasso I, Schmidt T, Heslop-Harrison JS (1995) Analysis of a repetitive DNA family from *Arabidopsis arenosa* and relationships between *Arabidopsis* species. Plant Mol Biol 27: 853–862

Kenton A, Parokonny AS, Gleba YY, Bennett MD (1993) Characterization of the *Nicotiana tabacum* L. genomes by molecular cytogenetics. Mol Gen Genet 240 : 159–169

Linde-Laursen IB, Ibsen E, Bothmer RV, Giese H (1992) Physical localization of active and inactive rRNA gene loci in *Hordeum marinum* ssp. *gussoneanum* (4 ×) by in situ hybridization. Genome 35 : 1032–1036

Mukai Y, Gill BS (1991) Detection of barley chromatin added to wheat by genomic in situ hybridization. Genome 34 : 448–452

Orozco-Castilho C, Chalmers KJ, Waugh R, Powell W (1994) Detection of genetic diversity and selective gene introgression in coffee using RAPD markers. Theor Appl Genet 87 : 934–940

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

Parokonny AS, Kenton AY (1995) Comparative physical mapping and evolution of the *Nicotiana tabacum* L. karyotype. In: Brandham PE, Bennett MD (eds) Kew Chromosome Conf IV. Royal Botanic Gardens, Richmond, Surrey, pp 301–320

Parokonny AS, Kenton AY, Gleba YY, Bennett MD (1994) The fate of recombinant chromosomes and genome interaction in *Nicotiana* asymmetric somatic hybrids and their sexual progeny. Theor Appl Genet 89 : 488–497

Raina SN, Narayan RKJ (1984) Changes in DNA composition in the evolution of *Vicia* species. Theor Appl Genet 68 : 187–192

Saghai Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc Natl Acad Sci USA 81 : 8014–8018

Schwarzacher T, Leitch AR, Bennett MD, Heslop-Harrison JS (1989) In situ localization of parental genomes in a wide hybrid. Ann Bot 64 : 315–324

Schwarzacher T, Anamthawat-Jonsson K, Harrison GE, Islam AKMR, Jia JZ, King IP, Leitch AR, Miller TE, Reader SM, Rogers WJ, Shi M, Heslop-Harrison JS (1992) Genomic in situ hybridization to identify alien chromosomes and chromosome segments in wheat. Theor Appl Genet 84 : 778–786

Thomas HM, Morgan WG, Meredith MR, Humphries MW, Thomas H, Leggett JM (1994) Identification of parental and recombined chromosomes in hybrid derivatives of *Lolium multiflorum* × *Festuca pratensis* by genome in situ hybridization. Theor Appl Genet 88 : 909–913

Wrigley G (1995) Coffee. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. Longman Scientific and Technical, UK, pp 438–443