Coffea Genome Structure and Relationship with Evolution

A. de KOCHKO[1], J. LOUARN[2], P. HAMON[1], S. HAMON[1], M. NOIROT[1]

[1]IRD, UR117, Centre IRD de Montpellier, BP 5045, 34032 Montpellier Cédex 1, France
[2]IRD station de Man, BP 434 Man, Côte d'Ivoire

SUMMARY

Among the numerous species constituting the *Coffea* genus, only one, *C. arabica* is tetraploid (2n=4X=44) all the others are diploid but their DNA content is quite variable. Originating from East Africa and Mascarene Islands, coffee trees are found all along the forest belt that crosses the continent. Interestingly, the genome of the species found closely to the origin center have the lowest amount of DNA and oppositely, the species that "migrated" have a much bigger genome in terms of DNA content. The main reason to explain this difference can be attributed to the nature and amount of repeated sequences. It has been shown that, as all the plants so far analyzed, coffee contains many different retrotransposons. It is assumed that the movement of some of these elements and the multiplication of copies in the genome can be related to the migration of the coffee trees and the stresses encountered. The augmentation of the DNA content also define new genetic barriers as the level of sterility in F1 hybrids issued from interspecific crosses is higher when the two parental species have bigger difference in DNA content. This sterility level is tightly related to the presence of univalent chromosomes during meiosis in the gamete precursor cells.

INTRODUCTION

Coffea genus comprises over 80 species and taxons all sharing essentially the same genome with a basic chromosome number of 11 (2x=2n=22). Only one species, *C. arabica* is tetraploid (2x=4n=44). It is generally accepted to consider that *Coffea* genus is originated from East Africa (region of North Kenya, Somalia, Ethiopia) and that it spread to West Africa on one side and to Madagascar and the Mascarene Islands on the other side. In each of these regions, secondary spreads took part, increasing the genetic variability of the genus and originating new species. A survey made on some species covering the major regions of the genus extent, showed that the nuclear DNA content is quite variable, it ranges from 0,95±0,13 pg to 1,78±0.33 pg for a diploid genome (2C value). *C. arabica* has of course a higher nuclear DNA content, with 2C=2.61±0.23 pg (Cros et al., 1995). This variation in DNA content can hardly be explained by an important interspecific difference in gene number. Even if such differences exist they can only account for a very low percentage of the total variation, instead, it was shown in many plants, but also in animals, that the major part of a genome is constituted by repetitive DNA which role is not always clearly defined.(Schmidt and Heslop-Harrison, 1998). This repetitive DNA is constituted by different kind of repeated sequences, among which: the telomeric and centromeric ones, necessary for chromosome stability and replication, various long or short repeats, satellites, microsatellites, tandems or palindromes which role needs still to be elucidated, transposons and retrotransposons which role in evolution has been shown (Lyubomirskaya and Ilyin, 1999) and the ribosomal RNA coding regions.

In coffee trees, interspecific hybrids can be obtained, but the more the DNA content varies between the parents the most sterile these hybrids are because of the increasing number of univalent chromosomes appearing during meiosis. Segregation distortions are also frequent

in interspecific hybrids, and the number of distorted alleles is more important when the difference in DNA content of the parents is most important.

It is very interesting to understand the origin of the diversity of *Coffea* genomes, identify the different repeated sequences present in each or all genomes, quantify and localize them. It could be possible to identify species specific, or even chromosome specific, sequences, follow the evolution of transposons and retrotransposons among the different species and relate their presence and frequency to the evolution of the genus.

RESULTS AND DISCUSSION

A *Sau* 3A genomic library enriched in repeated sequences was constructed using nuclear DNA from a F1 hybrid issued from a cross between *C. liberica* var Dewevrei and *C. pseudozanguebariae*. 193 clones were recovered and sequencing of the clones is underway. Up to date 36 sequences have been obtained and similitude searches (BLAST) have been conducted on data banks. The length of the sequenced fragment varies from 24 to 569 bp. The majority of the sequences so far analyzed (55%), showed no similarity with any data recorded in the banks. Very few coding sequences (5%) where detected showing significant similarity with known proteins or putative proteins. 33% of the fragments were similar to recognized repeated sequences including transposable elements which represented 33% of the repeated sequences (11% of the total). In Furthermore, the two major groups of retrotransposons, namely copia and Ty3-gypsy like, were found in our library. All these results are summarized in Table 1.

Table 1. Identification of some repeated sequences found in nuclear DNA of a F1 hybrid from a cross between *C. liberica* var Dewevrei and *C. pseudozanguebariae*

Seq #	length pb.	%AT	Comments, similitudes	Seq #	length pb.	%AT	Comments, similitudes
4	266	64	Copia-like, gag-pol polyprotein	15a	29	52	cp DNA rDNA spacer
6	569	58	H.s. ?	15b	132	62	?
7a	153	64	?	16b	43	67	?
7b	140	55	URF?	17a	116	65	Ty3/gypsy like, gag-pol polyprotein
8	469	61	?	17c	47	62	?
9	70	53	?	18	90	60	?
10	186	63	?	20	182	58	rDNA spacer
11a	53	60	?	21a	45	58	?
11b	185	53	Ty3/gypsy like, Integrase	21b	58	57	?
12	513	61	Rep. Seq.	21c	132	70	?
12a	339	62	?	21d	143	59	Protein CLB1 (L. esc.)
12b	185	60	Athila like?	33	563	61	Rep. Seq.
12c	24	67	?	35	265	68	L. esculentum Rep. Seq.
12d	35	57	?	36	527	63	
13a	136	68	?	38	444	65	ADNcp?
13b	25	60	?	40	466	60	H. sapiens, seq. Rep.
13c	206	80	?	54	448	61	Rep. Seq.
14	257	65	?	55	528	64	H. sapiens

Interestingly, it appears to be very promising to study the distribution and the organization of these retrotransposons in *Coffea* genomes and relate them to the spread and species differentiation within the genus. Indeed, it has been reported that the replication of retrotransposons might be induced by stress conditions (Grandbastien, 1998). The spread of an organism to a new environment is a stressful situation for that organism, retrotransposons might be activated under these conditions, in addition to the accumulation of mutations occurring during the retrotransposons insertion in new loci, the size of the genome will increase.

These events can be repeated several times on a period of time covering hundreds thousand years leading to new species with different coding and regulating capabilities and with genomes of very different sizes in terms of DNA content (not necessarily chromosomic numbers). This situation is encountered for the *Coffea* genus, as shown on Figure 1, the species with the lowest DNA content are found in East Africa, the alleged region of origin of the genus, and the species with the higher DNA content were identified is West Africa where they spread after a long migration across the African continent. Genome changes can also occur on much lower scale due to local variation in climatic conditions (Kalendar et al., 2000), which can explain, or at least be part of, secondary speciation.

It appears very important for the understanding of Coffea evolution to better characterize the nature and the amount of the retrotransposons present.

Figure 1. spreading of the *Coffea* genus in the African continent and the Indian Ocean slands

To identify species specific repeated sequences in order to follow introgressions. Survey the distribution of repeated sequences on the chromosomes to understand segregation distortions in interspecific progenies and try to estimate recombination possibilities. Design of new strategies for *Coffea* species sustainable improvement can largely benefit from this studies.

BIBLIOGRAPHY

Cros J, Combes MC, Chabrillange N, Duperray C, Monnot des Angles A, Hamon S (1995) Nuclear DNA content in the subgenus *Coffea* (*Rubiaceae*): inter- and intra-specific variation in African species. Canadian Journal of Botany **73**: 14-20

Grandbastien M (1998) Activation of plant retrotransposons under stress conditions. Trends in Plant Science **3**: 181-187

Kalendar R, Transkanen J, Immonen S, Nevo S, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proceedings of the National Academy of Sciences of the USA **97**: 6603-6607

Lyubomirskaya NV, Ilyin YV (1999) Eukaryotic mobile genetic elements: the past, the present, and the future. Molecular Biology **33**: 958-968

Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends in Plant Science **3**: 195-199