Crop Breeding and Applied Biotechnology 5:1-7, 2005
Brazilian Society of Plant Breeding. Printed in Brazil

Chemical diversity in coffee species of genebank of Instituto Agronômico do Estado de São Paulo

Adriano Tosoni da Eira Aguiar[1*], Luiz Carlos Fazuoli[1], Terezinha de Jesus Garcia Salva[1], and José Laércio Favarin[2]

Received 8 April 2005

Accepted 20 September 2005

ABSTRACT - *The objective of this work was to group different coffee species present in the genebank of the Instituto Agronômico by using some chemical variables. A total of thirty-nine plants belonging to seven species were analyzed for chemical seed components (soluble solids, lipids, trigonelline, chlorogenic acids and caffeine). The results evidenced that species could be grouped by lipids, chlorogenic acid, trigonelline and caffeine. The existence of great variation among and within species ranging from 24.12 to 30.65% for soluble solids; 8.68 to 17.49% for lipids; 0.32 to 2.15% for trigonelline; 2.91 to 6.38% for chlorogenic acid and 0.80 to 2.50% for caffeine was also observed, indicating the possibility to select plants of interest for the improvement of cultivated coffee species.*

INTRODUCTION

Coffee plants belong to the family *Rubiaceae* and genus *Coffea* and comprise approximately 100 identified taxa, which include all species of agronomical importance. The two commercially most important species are *C. arabica* and *C. canephora*, usually known as Arabica and Robusta coffee, respectively.

Coffee species are geographically distributed across the tropical area of Africa. The establishment of germplasm banks of the known species at a global level stationed in many countries is very important to include the wild forms, which can have extremely advantageous agronomic traits such as resistance and tolerance to diseases, nematodes, insects, drought, frost, and other biotic and abiotic factors, as well as differing plant traits (root system, stem, leaves, flowers, fruits, and seeds).

Coffee genebanks exist in a few countries although these collections have few accesses of species aside from *C. arabica* and *C. canephora*. At the Instituto Agronômico in Campinas (IAC), state of São Paulo, it was possible to collect 16 of the main species of the genus so far, besides different varieties of *C. arabica*, *C. canephora* and *C. liberica* species.

This plant material presents great genetic variability in several traits such as: stem traits and size of plant, leaves, fruits and seeds, resistance to biotic and abiotic factors, root system and cup quality. Furthermore, it forms a valuable genepool for different breeding purposes which does not only present differences among morphologic and

[1] Centro de Café "Alcides Carvalho", Instituto Agronômico (IAC/APTA), C. P. 28, 13.001-970, Campinas, SP, Brasil. *E-mail: aguiar@iac.sp.gov.br
[2] Departamento de Produção Vegetal, ESALQ/USP, C. P. l 9, 13.418-900, Piracicaba, SP, Brasil

agronomic traits but also at biochemical and molecular levels (Ky et al. 2001).

Numerous criteria have been used to determine genetic diversity of coffee genus such as some morphologic, agronomic, cytological, molecular, and geographic traits.

The objective of this research was to group different species present in the IAC Coffee genebank and to evaluate the existing variability by using chemical variables.

MATERIAL AND METHODS

Seeds of 39 plants from seven species were used. The plants were provided by the Coffee genebank of the Instituto Agronômico - IAC/APTA, at the Experimental Center of Campinas (Santa Elisa farm) in Campinas, SP.

Table 1 shows the relationship among species, their respective varieties and analyzed plant samples. The seven following species used in this investigation were: *C. canephora*, *C. liberica*, *C. congensis*, *C. eugenioides*, *C. stenophylla*, *C. racemosa*, and *C. kapakata*.

Mature fruits of each plant were collected individually at the berry stadium in 2002, according to the species' maturation time. The dried seeds were finely ground to perform the analyses.

The soluble solids content was determined using 10 grams of ground coffee according to methodology number 15.034 of AOAC (1997).

To estimate the total oil content, 10 grams of ground coffee were extracted with 100 mL petroleum ether in a Butt apparatus for 16 hours. The solid material was weighed after drying (30 minutes at 105 ºC). Lipid concentrations were calculated by the difference between the initial and degreased mass weight (Mazzafera et al. 1998).

Trigonelline, chlorogenic acid and caffeine were extracted in 70% methanol at 60 ºC for an hour (Mazzafera 1999) and quantified by High Performance Liquid Chromatography (HPLC) (Casal et al. 2000). The elution was performed in isocratic mode using a mobile phase 30%; 0.3%; 69.7%, v/v/v, methanol/ acetic acid/bidistilled water (pH 3.0) at a flow rate of 1mL min^{-1}. Trigonelline, chlorogenic acid and caffeine were quantified by comparing peak heights with standard values and the results presented as dry base.

The experiment comprised randomized complete block designs with two replications. Variance analyses were applied for the soluble solids, lipids, trigonelline, chlorogenic acid, and caffeine content. The Tukey test (P = 0.05) was applied to species means to identify statistically different species in relation to the evaluated traits. Multivariate analysis was used to determine the principal components (PCA), discriminant factors (DFA), the classification matrix, and Mahalanobis distance. All statistical analyses were performed on software STATISTICA (Guerrero and Suárez 2001).

RESULTS AND DISCUSSION

The results obtained for chemical variables (soluble solids, lipids, trigonelline, chlorogenic acid, and caffeine) for the seven species used in the present investigation are shown in Table 1.

Experimental coefficients of variation for soluble solids and trigonelline indicated a good precision ranging from 1.60 to 4.59%, respectively. Significant variation was observed for the contents of all analyzed chemical variables (Table 1). Variation was also observed among plants within species. These variations can be attributed to genetic differences among them.

The extreme values observed in the different species ranged from 24.12 to 30.65% for soluble solids; 8.68 to 17.49% for lipids; 0.32 to 2.15% for trigonelline; 2.91 to 6.38% for chlorogenic acid and 0.80 to 2.50% for caffeine. The soluble solids contents ranged from 24.12 to 30.35%. *C. canephora* (25.52% c) is a cultivated species with more soluble solids than *C. arabica*, another cultivated species. The trait high soluble solids is desirable for the soluble coffee industry. *C. liberica* var. dewevrei (29.02% a) and *C. congensis* (27.94% ab), both crossable with *C. arabica* (24.1%) have the potential to raise soluble solids contents to about 14% higher than in *C. canephora* var. *robusta*. There is additional phenotypic varibility available among *C. liberica* accesses, ranging from 26.22 to 30.46%. There are Arabica phenotypic selections with good agronomic traits and with *C. liberica* genes that can be explored to improve this trait in *C. arabica* selections such as the germplasm of Catuaí Sh_3 and Mundo Novo Sh_3.

Lipid contents are very important because of the high value of coffee oil on the market and the tocopherol content in the oil. Lipid contents in coffee species range

Table 1. Soluble solids, lipids, trigonelline, chlorogenic acid and caffeine content in coffee species

Species	Collection	Introduction number of plants	Soluble solids*	Lipids*	Trigonelline*	Chlorogenic acid*	Caffeine*
			g $100g^{-1}$				
C. canephora var.	5	1564	26.91	10.82	1.01	5.27	2.15
robusta	10	801	25.12	12.27	1.01	5.49	2.00
	12	784	24.53	9.64	0.97	5.68	2.22
	Mean**		25.52 c	10.91 cd	0.99 d	5.48 b	2.12 a
C. liberica var.	63	63	26.22	13.54	0.58	3.19	0.86
dewevrei	4	999	28.33	12.55	0.40	3.07	1.03
	5	766	30.46	12.39	0.32	3.71	0.98
	6	1000	29.67	14.50	0.43	3.02	0.93
	7	1001	27.41	15.11	0.39	3.15	0.87
	8	767	30.65	14.12	0.49	3.21	1.06
	9	751	29.57	12.95	0.61	3.98	1.13
	13	769	28.51	14.52	0.61	3.23	0.90
	15	798	30.35	14.07	0.65	3.08	0.92
	Mean**		29.02 a	13.75 b	0.50 e	3.29 e	0.96 d
C. congensis	11	4349	26.57	11.78	1.32	4.75	1.93
	12	4349	25.76	10.20	1.34	5.14	1.54
	14	4349	27.81	10.98	1.13	4.84	1.78
	15	4349	26.62	10.44	1.22	4.48	2.05
	83	4349	29.05	12.10	1.38	4.86	2.01
	84	4349	27.99	8.89	1.23	4.52	2.50
	85	4349	27.68	10.50	1.31	5.00	1.81
	86	4349	30.17	8.68	1.52	5.77	2.49
	87	4349	29.59	11.80	1.34	5.46	2.26
	106	4349	28.05	11.14	1.24	4.97	2.31
	107	4349	29.81	11.63	1.63	4.57	1.97
	110	4349	26.12	10.18	1.31	3.97	1.65
	Mean**		27.94 ab	10.69 d	1.33 c	4.86 c	2.03 a
C. eugenioides	1	1140-	24.13	16.68	1.85	5.27	0.95
	2	24	25.41	17.30	1.76	4.75	1.22
	3	1140-	25.79	16.59	1.70	4.57	0.82
	4	24	24.34	17.49	2.09	4.79	0.92
	8	1098-	24.12	16.69	1.96	4.73	0.80
	10	7	25.64	15.74	1.91	4.64	0.89
	Mean**	1098-	24.91 c	16.75 a	1.88 b	4.76	0.93 d
C. stenophylla	2	7	29.15	12.09	1.72	cd	1.59
	9	-	29.32	12.03	2.13	6.23	1.97
	10	2252	28.14	12.40	1.79	6.38	1.47
	Mean**		28.87 ab	12.17 c	1.88 b	6.05	1.68 b
C. racemosa	2[a]	1070-	24.93	9.73	1.45	6.22 a	1.49
	2b	13	25.64	13.29	1.31	5.10	1.53
	5[a]	1070-	28.23	12.50	1.15	5.41	1.17
	9[a]	13	27.55	11.54	1.44	3.97	1.07
	9c	1070-	29.84	9.54	1.29	4.77	1.08
	Mean**	13	27.24 bc	11.32 cd	1.33 c	2.91	1.27 c
C. kapakata	7		27.27 abc	16.05 ab	2.15 a	4.43 d	1.13 d
CV%			1.60	2.98	4.59		

* Mean of two replications

** Means in a column followed by the same letter are not significantly different according to Tukey's multiple range test at 5% probability

CV% = variation coefficients

from 10.69 to 16.75%. One of the ancestors of *C. Arabica*, *C. eugenioides*, presented a higher lipid content (16.75% a), about 35% more than *C. canephora* (10.91% cd). Some selection of *C. arabica* carrying *C. eugenioides* genes can therefore be applied to develop *C. arabica* cultivars with an increased oil content - dependent of the crossing, once the variability among *C. eugenioides* accessions ranges from 15.74 to 17.49%.

The genus *Coffea* has a high variability in trigonelline content, which is related to complex B vitamins. It is very important for health since it enriches coffee with vitamin B_3 (niacin) and B_7 (coline) besides other less important vitamins (Mazzafera 1991). *C. kapakata* has higher contents (2.15% a), 117% more than *C. canephora* (0.99% d) and will be very important to improve vitamin B contents in coffee. *C. eugenioides* (1.88% b) can be very important to improve vitamin B contents in *C. arabica* cultivars in the short term because it is crossable with *C. arabica* and contents ranged from 1.76 to 2.09% among accessions.

The bioflavonoid chlorogenic acid is very important as antioxidant (Amorin and Silva 1968) and ranged from 3.29 (*C. liberica*) to 6.22% (*C. stenophylla*). If breeding aims at reduced contents, the best source is *C. liberica* (3.02 to 3.98%), while *C. arabica* cultivars presented 3.2% (Martín et al. 1998). Four species that are easy to cross with *C. arabica* but can not be used to obtain a lower chlorogenic acid content in Arabica cultivars are *C. eugenioides* (4.76% cd), *C. congensis* (4.86% c), *C. racemosa* (4.43% d) and *C. canephora* (5.48 b).

Caffeine is the best-known component of coffee beverage and consumers generally prefer low caffeine contents or even caffeine-free coffee (Mazzafera et al. 1997). *C. arabica* cultivars have caffeine contents between 0.6 and 1.5% (Carvalho et al. 1965) and *C. canephora* between 1.50 and 3.50% (Ky et al. 2001). In this study the caffeine content ranged from 0.93 (*C. eugenioides*) to 2.12% (*C. canephora*). If breeding targets a low caffeine content, germplasm carrying *C. liberica* (0.86-1.13%) and *C. eugenioides* (0.80-1.22%) can contribute to improve the cultivated coffee species.

Based on the results of principal components analysis (PCA) species can be positioned and the chemical variables associated in a biplot (Figure 1). The first two principal components explained 79.3% of the total variation. In the biplot, the X axis characterized by the chlorogenic acid, caffeine and trigonelline accounted for 43.4% and Y axis, which was characterized essentially by the soluble solids and lipids, accounted for 35.9%. The variables chlorogenic acid and caffeine are highly correlated to the first principal component, demonstrating effectiveness at differentiating coffee species. Theses results show that the variables lipids, chlorogenic acid, trigonelline and caffeine contents can separate coffee species relative effectively, evidencing the existence of three species groups, as indicated in Figure 1.

Figure 1 shows that species on lower right (*C. eugenioides* and *C. kapakata*) formed one group and species on the upper left (*C. congensis*, *C. canephora*, *C. stenophylla* and *C. racemosa*) formed another. The species *C. liberica* on the upper right formed one more group, though a little distant from the other species. Our results agree well with earlier studies in literature, where the three identified groups were identified using chemical traits (Carvalho and Monaco 1967, Clifford et al. 1989, Anthony et al. 1993, Bridson 1994 and Mazzafera and Guerreiro Filho 1998). It should be emphasized that in an investigation realized by Mazzafera and Guerreiro Filho (1998), the authors worked with coffee pulp (a maternal tissue) to detect chemical components through HPLC, in contrast to our study.

In general, the results evidenced that most of the species from central and west Africa have superior caffeine contents (*C. canephora*, *C. congensis*, and *C. stenophylla*), with exception of *C. kapakata* and *C. liberica*, in contrast to eastern Africa species (*C. eugenioides* and *C. racemosa*). The low caffeine content in *C. kapakata* and *C. liberica* species could be explained by the following reason: in agreement with Anthony et al. (1993), based on observations of taxonomic criteria, *C. kapakata* has many affinities with species from east Africa. With regard to *C. liberica*, these differences between the present study and literature data are due to genetic variability present among coffee plants of this species. Is known that *C. liberica* is widely distributed across the African continent, extending from Guinea (west) until Zaire (central) (Dussert et al. 1999).

Table 2 presents the classification matrix and plant numbers of the different analyzed materials established through the discriminating factorial analysis (DFA). In the mean, about 97.4% of total plants concerning the different

species were classified in the respective species. In general, the existence of a low variability was observed in the classification within species among species (92.0 to 100.0%).

The plants of the *C. canephora*, *C. liberica*, *C. eugenioides*, *C. stenophylla*, *C. racemosa* and *C. kapakata* species were showed to be similar, indicating that all were classified in their proper species (Table 2). In an inferior position, although very well classified, is *C. congensis*. This species presented 92.0% of classification. It was observed that from the total of the *C. congensis* plants, only one was not classified as of the proper species, but as belonging to *C. racemosa*. These results confirm the fidelity of the groupings obtained by PCA.

The Mahalanobis' distances of the different analyzed coffee species were obtained (Table 3), whose values represent the divergence among the materials. The largest distance was verified between *C. kapakata* and *C. liberica* species (196.2), while *C. eugenioides* and *C. kapakata* presented the greatest similarity (12.4) among the analyzed materials. *C. canephora* and *C. congensis* are very similar, as well *C congensis* and *C. racemosa*, and *C. eugenioides* and *C. kapakata.*

On the other hand, *C. kapakata* is very different from *C. canephora*, *C. liberica* and *C. congensis*, as well as *C. stenophylla* from *C. liberica*, *C. eugenioides* from *C. canephora* and *C. liberica*, *C. congensis* from *C. eugenioides*, and *C. canephora* from *C.* kapakata. These results suggest that the distribution of biochemical diversity does not coincide with the species geographic distribution.

CONCLUSIONS

1. The variables lipids, chlorogenic acid, trigonelline and caffeine allowed the discrimination of the coffee species: (i) *C. congensis*, *C. canephora*, *C. stenophylla* and *C. racemosa*; (ii) *C. eugenioides* and *C. kapakata* and (iii) *C. liberica*.

2. The variable soluble solids was not effective at discriminating coffee species.

3. Results evidenced the existence of great variation among species and within species for all analyzed variables, indicating the possibility of selecting plants of interest to achieve enhanced cultivated species with higher trigonelline, lower caffeine and lower chlorogenic acid.

Table 2. Classification matrix and plants number among coffee species used in the present study

Species	Classification (%)	A	B	C	D	E	F	G
A	100.0	3	0	0	0	0	0	0
B	100.0	0	9	0	0	0	0	0
C	92.0	0	0	11	0	0	1	0
D	100.0	0	0	0	6	0	0	0
E	100.0	0	0	0	0	3	0	0
F	100.0	0	0	0	0	0	5	0
G	100.0	0	0	0	0	0	0	1
Total	97.4	3	9	11	6	3	6	1

A: *C. canephora*; B: *C. liberica*; C: *C. congensis*; D: *C. eugenioides*; E: *C. stenophylla*; F: *C. racemosa*; G: *C. kapakata*

Table 3. Estimates of the Mahalanobis distances among coffee species used in the present study

Species	*C. canephora*	*C. liberica*	*C. congensis*	*C. eugenioides*	*C. stenophylla*	*C. racemosa*	*C. kapakata*
C. canephora	-	67.2	12.9	120.6	67.8	32.3	156.7
C. liberica		-	75.7	151.6	137.4	56.8	196.2
C. congensis			-	82.3	35.7	14.1	96.4
C. Eugenioides				-	44.3	53.6	12.4
C. Stenophylla					-	27.5	57.3
C. racemosa						-	74.9
C. kapakata							-

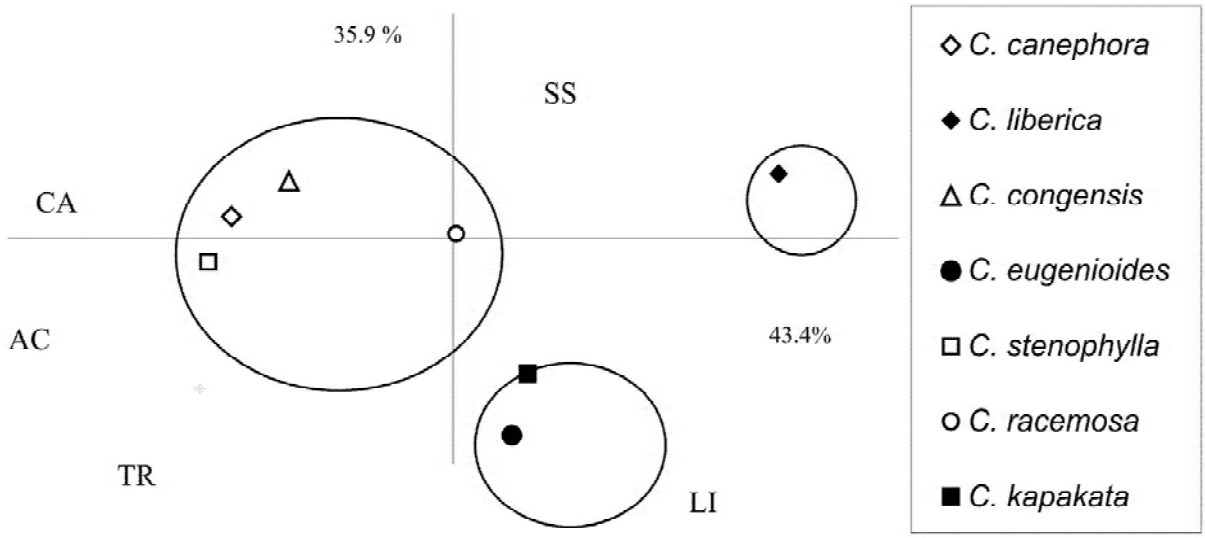

Figure 1. Display of principal component analysis related to seven *Coffee* species evaluated for soluble solids (SS), lipids (LI), caffeine (CA), trigonelline (TR) and chlorogenic acid (AC)

Diversidade química em espécies de café do banco de germoplasma do Instituto Agronômico do Estado de São Paulo

RESUMO - *Este trabalho teve por objetivo agrupar cafeeiros das espécies de Coffea presentes no Banco de Germoplasma de Café do Instituto Agronômico de Campinas, mediante a utilização de variáveis químicas dos grãos. Utilizaram-se um total de trinta e nove plantas pertencentes a sete espécies, que foram avaliadas em função das características químicas (sólidos solúveis, lipídios, trigonelina, ácido clorogênico e cafeína) de sementes. Os resultados evidenciaram que as variáveis*

lipídios, ácido clorogênico, trigonelina e cafeína mostraram-se eficientes no agrupamento das espécies. Observaram-se também a existência de grande variação entre e dentro dos diferentes materiais analisados, com valores extremos de 24,12 a 30,65% para sólidos solúveis; 8,68 a 17,49% para lipídios; 0,32 a 2,15% para trigonelina; 2,91 a 6,38% para ácido clorogênico e 0,80 a 2,50% para cafeína, indicando a possibilidade de selecionar plantas de interesse para o melhoramento de cafeeiros cultivadas.

Palavras-chave: variabilidade genética, melhoramento, espécies de café, qualidade.

REFERENCES

Amorin HV and Silva OM (1968) Relationship between the polyphenoloxidase activity of coffee beans and the quality of the beverage. **Nature 219**: 381-382.

Anthony F, Clifford MN and Noirot M (1993) Biochemical diversity in the genus *Coffea* L.: chlorogenic acids, caffeine and mozambioside contents. **Genetic Resources and Crop Evolution 40**: 61-70.

Association of analytical chemists (1997) **Official methods of analysis**. Washington, 726p.

Bridson DM (1994) Additional notes on *Coffea* (Rubiaceae) from Tropical East Africa. **Kew Bulletin 49**: 331-342.

Carvalho A and Mônaco LC (1967) Genetic relationships of selected *Coffea* species. **Ciência e cultura 19**: 151-165.

Carvalho A, Tango JS and Mônaco LC (1965) Genetic control of caffeine content in coffee. **Nature 205**: 314.

Casal S, Oliveira B and Ferreira MA (2000) HPLC/diode-array applied to the thermal degradation of trigonelline, nicotinic acid and caffeine in coffee. **Food Chemistry 68**: 481-485.

Clifford M N, Williams T and Bridson D (1989) Chlorogenic acid and caffeine as possible taxonomic criteria in *Coffea* and *Psilanthus*. **Phytochemistry 28**: 829-838.

Dussert S, Lashermes P, Anthony F, Montagnon C, Trouslot P, Combes MC, Berthaud J, Noirot M and Hamon S (1999) Le caféier, *Coffea canephora*. In: Hamon P, Seguin M, Perrier X and Glaszmann J C (eds.) **Diversité génétique des plantes tropicales cultivées**. CIRAD, Montpellier, p.175-194.

Guerrero G and Suárez M (2001) Chlorogenic acids as a potencial criterion in *Coffee* genotype selections. **Food Chemistry 49**: 2454-2458.

Ky CL, Louarn J, Dussert S, Guyot B, Hamon S and Noirot M (2001) Caffeine, trigonelline, chlorogenic acid and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. **Food Chemistry 75**: 223-230.

Martín JM, Pablos F and González AG (1998) Discrimination between arabica and robusta green coffee varieties according to their chemical composition. **Talanta 46**: 1259-1267.

Mazzafera P (1991) Trigonelline in coffee. **Phytochemistry 30**: 2309-2310.

Mazzafera P (1999) Chemical composition of defective coffee beans. **Food Chemistry 64**: 547-554.

Mazzafera P and Guerreiro Filho O (1998) Ultraviolet HPLC-derived profiles as a tool in *Coffea* (Rubiaceae) taxonomy. **Journal of Comparative Biology 3**: 15-20.

Mazzafera P, Silvarolla MB, Lima MMA and Medina Filho HP (1997) Caffeine content of diploid coffee species. **Ciência e Cultura 49**: 216-218.

Mazzafera P, Soave D, Zullo MAT and Guerreiro Filho O (1998) Oil content of green beans from some *Coffee* species. **Bragantia 57**: 45-48.