Legume Res., **33** (4) : 242 - 248, 2010

AGRICULTURAL RESEARCH COMMUNICATION CENTRE
www.arccjournals.com / indianjournals.com

CLASSIFICATION OF NEW GERMPLASM AND ADVANCED BREEDING LINES OF GROUNDNUT (*ARACHIS HYPOGAEA* L.) THROUGH PRINCIPAL COMPONENT ANALYSIS

I. Sudhir Kumar*, N. Marappa and M. Govindaraj

AICRP on Groundnut, Agricultural Research Station,
Chintamani - 563 125, Karnataka, India.

ABSTRACT

Sixty four accessions of groundnut (*Arachis hypogaea* L.) evaluated under late *kharif* season have been classified using Principal component analysis (PCA) based on correlation matrix yielding eigen values and eigen vectors. Fourteen principal components (PC) have been extracted using the mean performance of the genotypes, first ten principal components contributed over 95% of variation. Relative positive weights by each of the component to each single character has shown pod yield per plant and kernel yield per plant being given high positive weight by first principal component. Biplot of first two principal components showed characters *viz.*, plant height, harvest index, pod yield, kernel yield per plant and oil yield per plant distinguishing among the accessions along the first principal component vector. Cluster analysis was performed based on first ten PC scores accounting more than 95% of variation which classified the sixty four accessions in to three clusters. Accessions in clusters 1 and 2 showed a wide range for several agronomic characters. This provides convenience in selecting superior accessions from each of these clusters for various yield contributing in the future breeding programs.

Key words : Advanced breeding lines, Principal component analysis, Clusters, Groundnut.

INTRODUCTION

Groundnut (*Arachis hypogaea* L.) is an annual legume and is the premier oilseed crop of India with potential to produce high edible oil and easily digestible proteins in its seeds. Groundnut oil contains a higher proportion of unsaturated fatty acids, including essential fatty acids like linolenic and linoleic acids (Desai *et al.,* 1999). In Southern Karnataka, groundnut being the principal oil seed crop was grown in an area of 5.5 lakh ha during *kharif* under rainfed condition with an average productivity of 650 kg/ha (Anonymons, 2006). The area under this crop is decreasing due to delayed onset of monsoons in this zone. Hence there is need to incorporate breeding lines suitable for late *kharif* season from the existing germplasm. Studies on available germplasm under the environment, where it is to be exploited is essential for successful utilization of germplasm resources for the development of superior varieties.

For utilizing the most potential, recapitulation of genetics of the quality and the yield contributing characters would be of great importance for the cause of plant improvement. Proper exploitation of the available variability in crop with the primary objective to identify and select superior genotypes with desirable trait from a broad array of breeding material is foremost. Principal component analysis is a variability estimation method that reduces data dimensionality by performing covariance analysis between factors. PCA reduces the number of variables by finding new components that are

*Present address: Sorghum Breeding, ICRISAT, Patancheru 502 324, India.

combinations of the old variables. It determines the net effect of each variable on the total variance of the data set and then extracts the maximum variance possible from the given material. In the present study the mean measure of sixty four accessions of groundnut were subjected for principal component analysis as variability estimation through classification of genotypes.

MATERIAL AND METHODS

The material for the present investigation comprised of 64 genotypes (34 new germplasm accessions, 24 advanced breeding lines and 6 controls) of groundnut grown in 8 × 8 simple lattice design with two replications during *kharif* 2006 (August) in the field unit of All India Coordinated Research Project (AICRP) on Groundnut, Agricultural Research Station, Chintamani, representing Southern Karnataka. Observations were recorded on ten randomly selected plants in each genotype within the replication for fourteen characters, *viz.*, plant height, branches per plant, days to 50% flowering, days to maturity, matured pods per plant, pod yield per plant, kernel yield per plant, shelling percentage, 100-kernel weight, sound mature kernel percent, harvest index, oil content, oil yield per plant and specific leaf area. The oil content in per cent was determined directly using nuclear magnetic resonance spectrometer (NMR) unit at Dept. of Oil Seeds, TNAU, Coimbatore. The specific leaf area (SLA) was recorded at 70 DAS utilizing the 3rd or 4th leaf from the top on the main axis. Leaf area of each leaf was measured by an automatic leaf area meter. Dry weight of leaves was recorded after oven drying at 80^{0}C for 48 hrs and SLA was calculated using the formula SLA (cm2/g) = Leaf area (cm2) / Oven dry weight of leaf (g). Principal component analysis (PCA) was carried out using XLSTAT (version 2010, 2.02) with mean observations of all the traits and Cluster analysis according to Ward (1963) was performed using the scores of first ten principal components resulting in three clusters in NTSYS software (Rohlf, 1990).

RESULTS AND DISCUSSION

The principal component analysis of sixty four accessions based on correlation matrix yielded the eigen roots (eigen values) and eigen vectors. The eigen values and associated cumulative percentage of variation explained by eigen vectors have been presented in Table 1 and Fig.1 which shows the cattle scree graph for variation explained by various principal components. The eigen root of first principal component accounted for 28.15% of total variation presented in the original data followed by second to sixth principal component which accounted for 18.89, 12.49, 8.92, 6.90, 6.47, and 5.11 per cent respectively. The first six principal components accounted for 81.83% of total variation whereas, the percentage of variation explained for seven to fourteen PCs was small (18.16). All components together explained more than 95% of variation of the original data units.

The eigen vector of the fourteen principal components has been scaled in such a way that the largest element in each vector as unity was interpreted as the relative weight given to the variable in each component. The important variables are those which have high negative and positive weight values among all the vectors which were utilized for ordination. It is an ordination method based on variance and covariance of an element based on vector. The basic subject of principal component analysis was to construct new variables from the original character in an optimizing way so that the information lost during formation of new variables is kept to minimum. This provides a multidimensional view of variation in a multivarietal setup. The principal components with 95% of total variance were taken for summarization of original data on groundnut germplasm in reduced dimension according to the approach of Rao (1964) which based on covering 90% of total variables had seen to be more appropriate and has been adopted by most of the workers. The scales per corresponding eigen vector for principal component taking the largest element

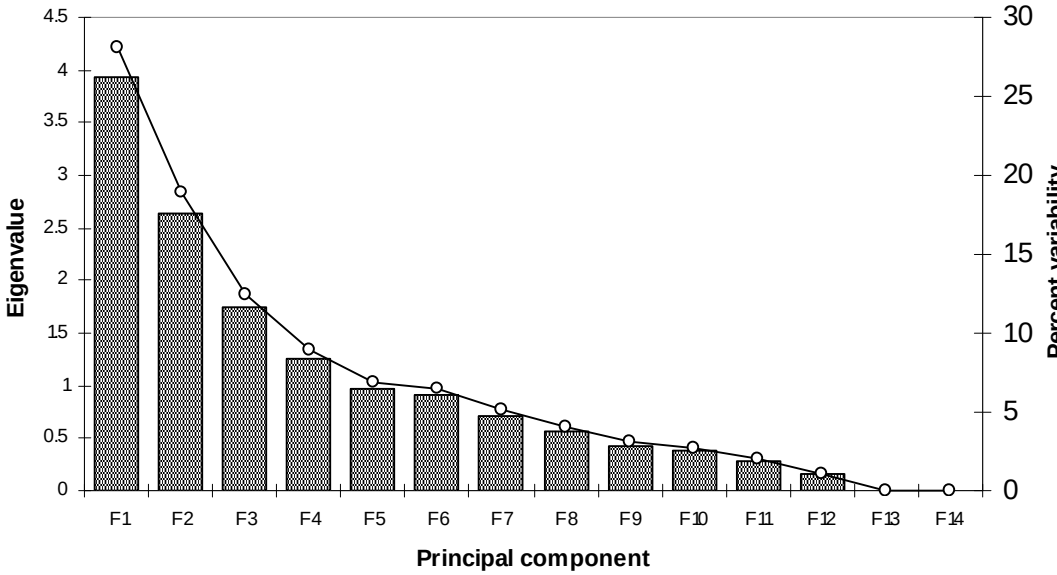

Fig 1: Cattle Scree graph for variation explained by various principal components on groundnut germplasm.

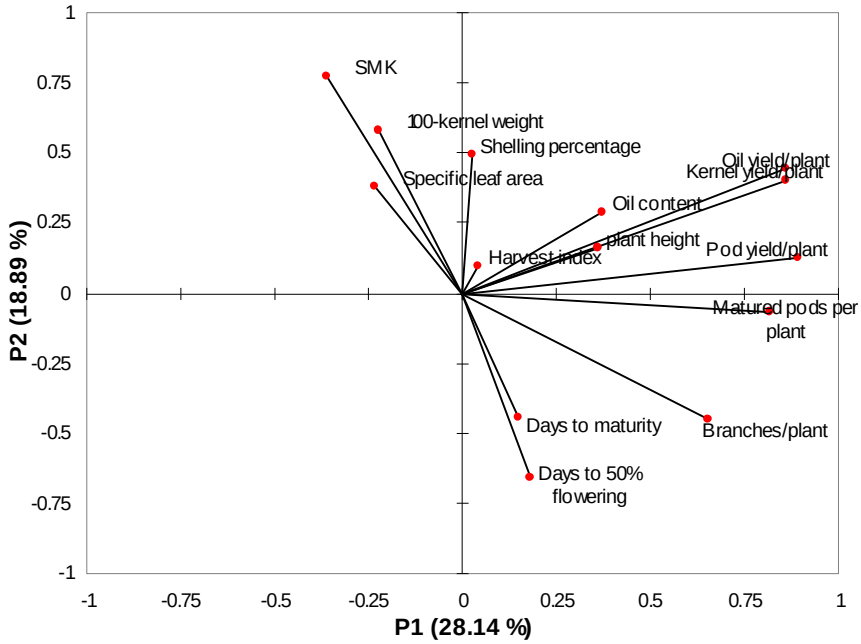

Fig 2: Biplot of the fourteen quantitative characters along the first and second principal component vectors based on sixty four accessions of groundnut.

in each vector as unity has been presented in Table 1. Jeffers (1967) suggested that these elements may be interpreted as the relative weight given to the variable in each component and important variable all those which have high positive and negative weight. The first principal component gave high positive weight (0.44) to pod yield per plant and kernel yield per plant, similarly second, third, fourth and fifth Principal components gave high positive weights to sound mature kernel percent (0.47), shelling percentage (0.53), specific leaf area (0.65) and oil content (0.59) respectively. The principal component analysis summarized large multidimensional data into a reduced dimension in

Table 1: Eigen vector, eigen root and associated variation for different principal components in Groundnut germplasm

Character	Eigen vectors													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Plant height (cm)	0.18	0.10	-0.43	-0.33	0.28	0.38	0.06	0.28	0.17	0.57	0.06	0.09	-0.02	-0.01
Branches per plant	0.33	-0.28	0.02	0.02	0.23	-0.20	-0.17	0.16	0.58	-0.31	0.47	0.11	-0.02	0.00
Days to 50% flowering	0.09	-0.40	0.21	0.16	0.25	0.04	-0.43	-0.49	-0.15	0.45	0.02	0.22	-0.01	-0.01
Days to maturity	0.08	-0.27	0.14	0.51	-0.04	0.43	-0.18	0.60	-0.14	-0.05	-0.15	-0.08	0.00	0.00
Matured pods per plant	0.41	-0.04	-0.22	0.13	-0.21	-0.19	0.20	0.04	-0.11	-0.08	-0.32	0.72	0.01	-0.01
Pod yield per plant (g)	0.45	0.08	-0.17	0.04	-0.28	0.06	-0.13	-0.14	-0.14	0.00	0.17	-0.29	0.67	-0.24
Kernel yield per plant (g)	0.44	0.25	0.12	-0.05	-0.13	-0.08	-0.21	0.00	-0.05	0.07	-0.06	-0.21	-0.19	0.76
Shelling percentage	0.01	0.31	0.53	-0.16	0.23	-0.24	-0.16	0.26	0.23	0.15	-0.39	0.10	0.38	-0.14
100 Kernel weight (g)	-0.11	0.36	-0.28	0.18	0.09	0.38	-0.40	-0.31	0.35	-0.32	-0.31	0.10	0.00	0.00
Sound mature kernel %	-0.18	0.47	0.01	0.08	0.06	-0.04	-0.31	0.19	-0.39	-0.03	0.55	0.38	0.02	0.00
Harvest index (%)	0.02	0.06	0.49	-0.13	-0.47	0.51	0.20	-0.14	0.27	0.07	0.23	0.26	-0.02	0.01
Oil content (%)	0.19	0.18	0.17	0.28	0.59	0.22	0.54	-0.20	-0.14	-0.15	0.09	0.00	0.17	0.15
Oil yield per plant (g)	0.43	0.27	0.15	0.03	0.04	0.00	-0.05	-0.05	-0.09	0.00	-0.03	-0.19	-0.58	-0.57
Specific leaf area (cm2/g)	-0.12	0.23	-0.09	0.65	-0.18	-0.27	0.17	-0.03	0.37	0.46	0.11	-0.08	-0.03	-0.01
Eigen Root	0.18	0.10	-0.43	-0.33	0.28	0.38	0.06	0.28	0.17	0.57	0.06	0.09	-0.02	-0.01
Variation (%)	28.145	18.893	12.499	8.922	6.905	6.475	5.112	4.078	3.054	2.709	2.040	1.109	0.056	0.004

the form of orthogonal component. Also an attempt has been made to observe the variation explained by fourteen quantitative characters along one and two principal component vectors (Fig. 2) whose variation accounted for 25.89% and 16.21% respectively. Characters distinguishing within the given set of germplasm along the first principal component vector are plant height, branches per plant, harvest index, pod yield, matured pods per plant, kernel yield per plant and oil yield per plant which showed major contribution of these characters towards variation. It can be said that the selection of genotypes for transgressive breeding may be based on these above characters whose contribution of variation was maximum.

Ward's hierarchical cluster analysis based on first ten principal component scores (total variation accounted more than 90%) resulted in three clusters (Fig. 3). The first cluster comprised thirty six accessions which include fourteen new germplasm lines and four advanced breeding lines. The second cluster comprised eighteen accessions of which four were advanced breeding lines rest of them were new germplasm lines. Third cluster comprised ten accessions including six control varieties (JL 24, TMV 2, Narayani, CTMG 1, VRI 2, GPBD 4) and four new germplasm lines. Upadhyaya *et al.* (2006) evaluated a core collection of 21 early maturing landraces deriving three clusters based on principal component scores in which 2 and 3 clusters showed wide range for several agronomic characters. In the present study involving germplasm from different geographical regions, no distinct relationship was observed between geographical origin and cluster formation. The range, mean and variance of each character in all three clusters are presented in Table 2. Cluster three was found to have higher mean values for 100-kernel weight, sound mature kernel percent and specific leaf area while cluster two was found to have high mean values for the remaining characters under study. Clusters one and two were found to have wide range for all the characters under study. Hybridization between genotypes of different

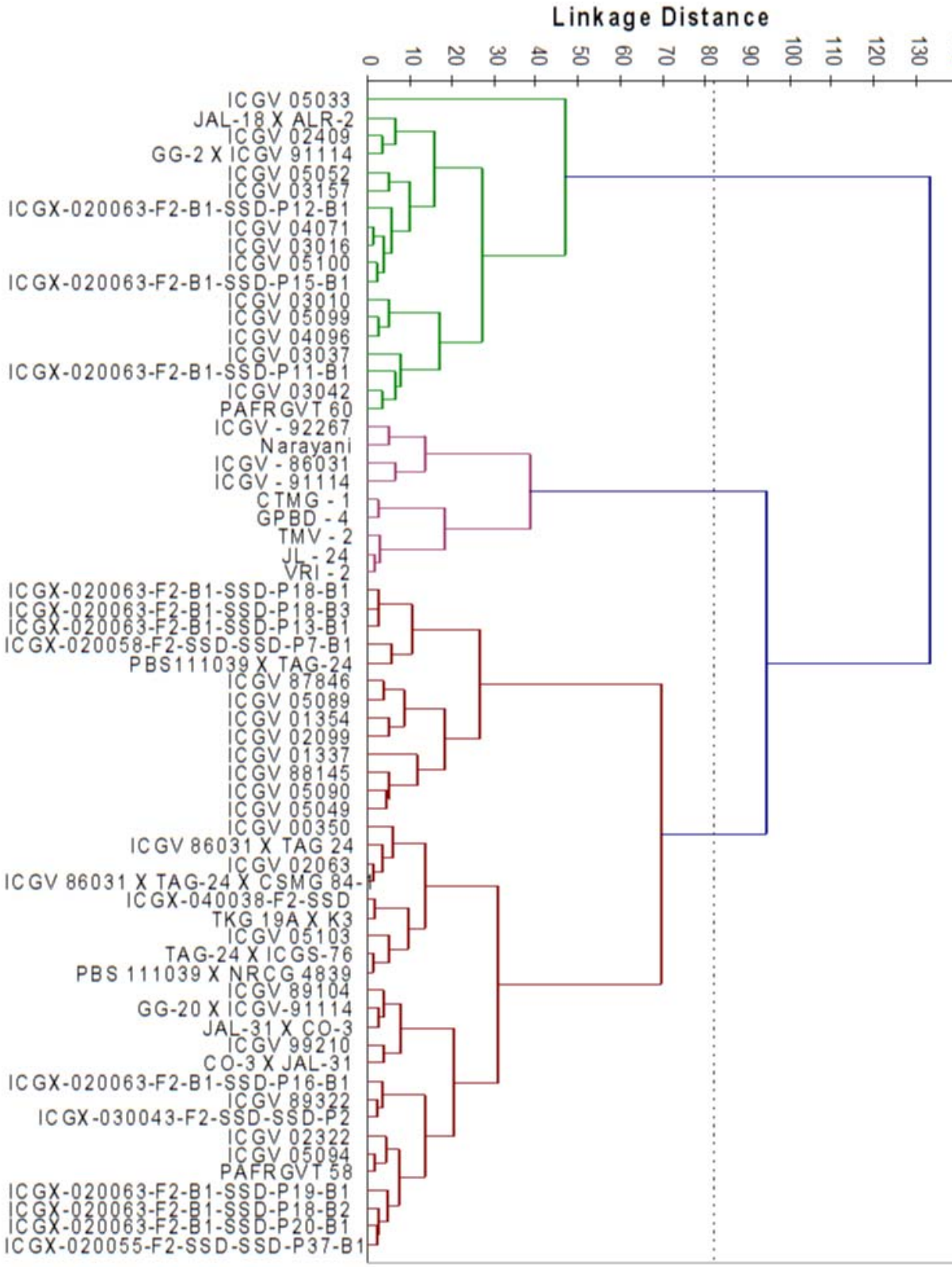

Fig. 3: Dendrogram showing sixty four accessions of groundnut based on the scores of first ten principal components.

Table 2: Range, mean and variance of three clusters of groundnut germplasm.

Character	Range			Mean			Variance		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
Plant height (cm)	19.20-34.15	20.00-34.00	20.15-34.80	26.28	28.23	26.40	0.14	0.13	0.17
Branches / plant	4.90-9.50	6.10-9.30	4.10-8.00	7.37	8.19	5.50	0.18	0.13	0.22
Days to 50% flowering	33.50-40.50	36.50-39.50	30.50-39.50	38.08	37.97	35.10	0.05	0.02	0.08
Days to maturity	100.50-121.00	102.50-121.50	104.50-108.00	111.89	112.36	110.35	0.05	0.04	0.04
Matured pods / plant	12.50-23.00	14.00-23.00	13.50-16.50	15.72	18.50	14.80	0.17	0.13	0.06
Pod yield / plant (g)	12.00-20.90	12.50-24.50	11.75-18.10	14.37	16.73	13.68	0.13	0.20	0.13
Kernel yield / plant (g)	7.15-14.10	9.05-19.85	7.65-13.30	10.28	12.00	9.78	0.15	0.19	0.18
Shelling percentage	58.24-81.70	59.14-81.05	59.21-80.57	71.76	72.09	71.49	0.11	0.08	0.11
100 Kernel weight (g)	19.45-37.10	20.20-34.45	26.05-38.00	27.48	27.13	34.41	0.16	0.14	0.10
Sound mature kernel %	57.00-79.00	59.50-75.50	75.00-87.50	68.92	67.22	79.70	0.08	0.06	0.05
Harvest index (%)	22.05-42.20	21.02-41.50	25.50-34.00	32.59	32.81	30.06	0.17	0.14	0.09
Oil content (%)	36.00-49.15	42.85-50.00	40.05-46.60	44.37	45.81	44.48	0.06	0.05	0.05
Oil yield / plant (g)	3.29-6.83	4.08-9.69	3.19-5.94	4.57	5.51	4.36	0.17	0.22	0.20
Specific leaf area (cm^2/g)	83.85-147.95	93.70-134.95	120.35-187.75	115.19	116.50	148.15	0.11	0.10	0.14

clusters with high cluster mean will result into transgressive segregants with high yield potential (Sisodia *et al.*, 1983). Therefore the genotypes from clusters one and two which showed high range and mean for important characters like pod yield, kernel yield, sound mature kernel percent, oil content can be used in the future breeding programme to recombine these traits. The present investigation reflects the importance of principal component analysis as variation reduction method and informative intensive method to examine the extreme genotypes while evaluating a large number of genotypes in particular.

REFERENCES

Anonymous (2006) Directorate of Economics and Statistics, Statistical Tables, KSDA, Bangalore, 1-10pp.

Desai, B.B. Kotecha, P.M. and Salunkhe, D.K. (1999) Science and Technology of Groundnut, Biology, Production, Processing and Utilization. Kalyani Publishers, New Delhi, 5pp.

Jeffers J.N.R. (1967) *Applied Statistics* **16**: 225-236.

Rao, C.R. (1964) Advanced Statistical Methods in Biometrics Research. Hafner Pub. Co. Darion 348-371.

Rohlf, R.J. (1990). NTSY S-PC, Numerical taxonomy and multivariate analysis system, version 1.8, Exeter Software, Setauket, New York.

Sisodia, N.S. Henry A. and Gupta, Y.K. (1983) *Madras Agricultural Journal* **70**(10): 678-680.

Upadhyaya, H.D. Reddy, L.J. Gowda, C.L.L. and Sube Singh. (2006) *Field Crops Research* **97**: 261-271.

Ward, J. (1963) Heirarchial grouping to optimize an objective function. *Journal of American Statistical Association* **38**: 236-234.

XLSTAT, Version 2.02, 2010 Addinsoft SARL, Paris, FRANCE.