Phylogenetic Analysis of the *WRKY* Transcription Factors Gene Superfamily in Coffee Plants

D. FERNANDEZ[1], D. RAMIRO[1,2], A-S. PETITOT[1], M. MALUF[3]

[1]IRD (Institut de Recherche pour le Développement), UMR1097 IRD/CIRAD/INRA, BP64501, 34032 Montpellier Cedex 5, France
[2]IAC (Instituto Agronômico de Campinas), Centro de Café 'Alcides Carvalho' – CP 28, 13.001-970 Campinas, SP – Brasil
[3]Embrapa Café, Centro de Café 'Alcides Carvalho' – CP 28, 13.001-970 Campinas, SP – Brasil

SUMMARY

WRKY family proteins are transcription factors involved in the regulation of development and plant defense response pathways. The *Arabidopsis thaliana* WRKY superfamily is made of 75 members. Common to these proteins is a DNA-binding region of approximately 60 amino acids in length which comprises the absolutely conserved sequence motif WRKY adjacent to a novel zinc-finger motif. A comparative phylogenetic analysis of the *WRKY* gene family in coffee and *A. thaliana* was conducted to assess the diversity of this family in coffee and to identify homologous coffee genes with putative function in defense responses to pathogens. Bioinformatic analysis of around 200 000 coffee Expressed Sequence Tags (ESTs) identified 313 ESTs with BLAST homologies to WRKY proteins. Almost 30 different putative WRKY genes were obtained, but only 25 unigenes encoding a protein with a WRKY domain were identified. Alignement of the WRKY domain sequences of the 25 coffee unigenes together with those of 72 *A. thaliana WRKY* genes showed a high conservation of the WRKY motif and the zinc-finger motif in the coffee WRKY domain. The 25 coffee WRKY members were distributed among the 3 main *A. thaliana* WRKY subgroups, with group I members displaying two WRKY domains, as expected. Conservation of the intron position within the WRKY domain sequence was evidenced when cloning the genomic sequence of one *WRKY* coffee gene (*CaWRKY1*). Clustering of the coffee *WRKY* genes based on the EST distribution in cDNA libraries made from tissues under several physiological conditions allowed to identify genes associated with development or with plant defense responses. To assess the involvement of *WRKY* genes in the coffee defense response pathways, gene expression patterns are being tested in coffee plants under several defense-related conditions.

INTRODUCTION

WRKY proteins are plant transcription factors encoded by a multigene family comprising over 74 genes in *Arabidopsis thaliana* (Eulgem et al., 2000 ; Dong et al., 2003) and more than 80 in rice (*Oriza sativa*) (Xie et al., 2005). WRKY proteins are characterized by the presence of one or two DNA - binding domains which comprise the conserved WRKYGQK core motif (Eulgem et al., 2000). Transcriptional regulation of a number of genes involved in several physiological processes may be driven by WRKY transcription factors (Eulgem et al., 1999; Ülker and Somssich, 2004). So far, recent studies have shown that WRKY proteins probably have regulatory functions in seed development, sugar signalisation and plant defence responses to pathogens (for review Ülker and Somssich, 2004). Indeed, pathogen infection, wounding or treatment with salicylic acid (SA) have been shown to induce rapid expression

of several *WRKY* genes from a number of plants (Dong et al., 2003; Ryu et al., 2006). In coffee (*Coffea arabica*), the *CaWRKY1* gene displayed altered expression patterns in response to biotic and abiotic treatments (Fernandez et al., 2004; Ganesh et al., 2006). Identification of regulatory genes involved in several physiological mechanisms such as disease resistance or seed development would offer new tools for improving coffee (*C. arabica*) varieties for important agronomic traits. The aim of this study was to identify *WRKY* genes in the coffee genome by data mining large sets of Expressed Sequence Tags (ESTs) and to predict their involvement in different physiological processes based on their expression patterns.

RESULTS

Identification of coffee *WRKY* genes

Table 1. List of coffee unigenes encoding a putative WRKY transcription factor.

Coffee clone	Origin	AtWRKY best BlastX	*A. thaliana* group	Expression group
CaWRKY-C5	*C. arabica*	33	I	C
CaWRKY-FR2-5E8	*C. arabica*	33	I	A
CaWRKY-126831	*C. Canephora*	33	I	Pericarp
CaWRKY-C10	*C. arabica*	33	I	A
CaWRKY-23-A03	*C. arabica*	44	I	Rust-induced
CaWRKY-119460	*C. Canephora*	40	IIa	Early-stage cherry
CaWRKY-C14	*C. arabica*	40	IIa	C
CaWRKY-130063	*C. arabica*	40	IIa	Early-stage cherry
CaWRKY-C23	*C. arabica*	40	IIa	A
CaWRKY1	*C. arabica*	460	IIb	Rust-induced
CaWRKY-C2	*C. arabica*	31	IIb	C
CaWRKY-C4	*C. arabica*	57	IIc	B
CaWRKY-C18	*C. arabica*	75	IIc	C
CaWRKY-C22	*C. arabica*	21	IId	C
CaWRKY-FR2-82A10	*C. arabica*	74	IId	A
CaWRKY-130733	*C. Canephora*	21	IId	Early-stage cherry
CaWRKY-125957	*C. Canephora*	15	IId	Pericarp
CaWRKY-C25	*C. arabica*	7	IId	C
CaWRKY-CB1-73G5	*C. arabica*	11	IId	B
CaWRKY-C24	*C. arabica*	27	IIe	A
CaWRKY-EA1-7B7	*C. arabica*	14	IIe	A
CaWRKY-125811	*C. Canephora*	69	IIe	Leaf
CaWRKY-C12	*C. arabica*	53	III	A
CaWRKY-C13	*C. arabica*	53	III	A
CaWRKY-C21	*C. arabica*	70	III	B
CaWRKY-C28	*C. arabica*	54	III	B

Coffee *WRKY* genes were retrieved from ESTs databases by keyword searches of annotated unigenes as well as by multiple BLAST searches using the WRKY domain sequence. The databases searched included (i) the Brazilian Coffee Genome Project ESTs database (http://www.lge.ibi.unicamp.br) which comprises more than 30 000 unigenes isolated from 27

cDNA libraries made from coffee (mostly *C. arabica*) tissues under several physiological conditions (Vieira et al., 2006), (ii) the *C. canephora* ESTs database developed from 5 cDNA libraries made from coffee leaves and seeds at a range of developmental stages (http://www.sgn.cornell.edu) and comprising more than 13 000 unigenes (Lin et al., 2005) and (iii) the IRD *C. arabica* EST database made of 1900 unigenes from defence-specific subtractive cDNA libraries (Fernandez et al., 2004; Lecouls et al., 2006).

We identified 313 ESTs with BLAST homologies to WRKY proteins. Search for the specific DNA-binding protein domain (WRKYGQK sequence followed by a C2H2- or C2HC-type of zinc finger motif) (Eulgem et al., 2000) was manually performed on the coffee unigene sequences. Almost 30 different putative *WRKY* genes were obtained, but only 25 unigenes encoding a protein with one or two WRKY domains were identified (Table 1). The remaining unigene sequences either did not cover the WRKY domain or ended within the domain, thus impairing further analyses.

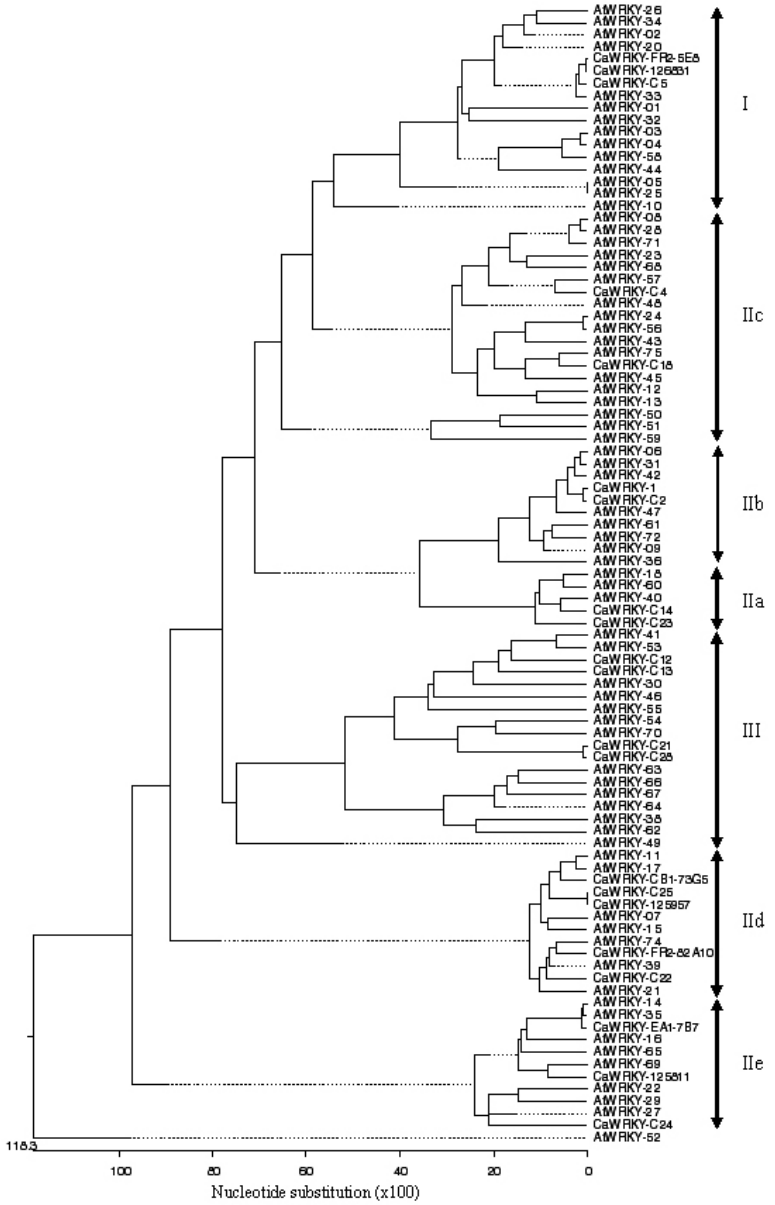

Figure 1. Dendrogram showing phylogenetic relationships between coffee and *A. thaliana* WRKY domains. Numbers on the right are the phylogenetic groups assigned to *A. thaliana* WRKY proteins (Eulgem et al., 2000).

Classification of *WRKY* genes on the basis of the WRKY domain sequences

BLAST homology to *A. thaliana* WRKY sequences were searched in GenBank database. The C-terminal WRKY domain sequences (68 amino acid residues) of 72 *A. thaliana WRKY* genes and the 25 coffee unigenes were aligned and a phylogenetic tree was constructed using the Lasergene software package (DNAStar, Inc., USA). Coffee genes were classified into the 3 main *A. thaliana WRKY* genes groups (Eulgem et al., 2000) (Figure 1 and Table 1). A high conservation of the WRKY motif and the zinc-finger motif was observed between the two plants. Group 3 *WRKY* coffee genes had a C2HC-type zinc-finger motif (C-$(X)_7$-C-$(X)_{23}$-H-*X*-C) whereas all other coffee WRKY genes had a C2H2-type (C-$(X)_n$-C-$(X)_p$-H-X-H).

Alignement of the *CaWRKY1* genomic and cDNA sequences (Petitot et al., 2006) showed the presence of an intron within the WRKY domain. The intron position (after the first Q residue of the zinc-finger domain) was highly conserved with that of *A. thaliana WRKY* genes (Eulgem et al., 2000).

Hierarchical classification of ESTs into expression groups

To identify coffee *WRKY* genes putatively associated with important physiological mechanisms such as development or plant defense responses, we analyzed the distribution of 17 *C. arabica WRKY* unigenes into the 27 cDNA libraries of the Brazilian coffee genes database. The presence/absence of WRKY ESTs in each cDNA library was recorded as a (0;1) matrix and used to construct a distance matrix (Simple-matching index) and a dendrogram with the UPGMA algorithm (Sneath and Sokal, 1973) contained in the software package TREECON, version 1.3b (Van de Peer and De Wachter, 1994). Coffee unigenes could be separated into 3 main groups based on their library distribution (Figure 2). The first cluster (expression group A) grouped unigenes only present in cDNA libraries involved in plant development (different fruit stages, embryogenic calli and lines), the second cluster (expression group B) contained ESTs from a cDNA library made from acibenzolar-S-methyl and brassinosteroide-induced tissues. The remaining unigenes (expression group C) were each largely distributed over 4-10 cDNA libraries and could not be assigned to a particular physiological trait. Future work will aim at identifying coffee *WRKY* genes involvement in agronomically important traits.

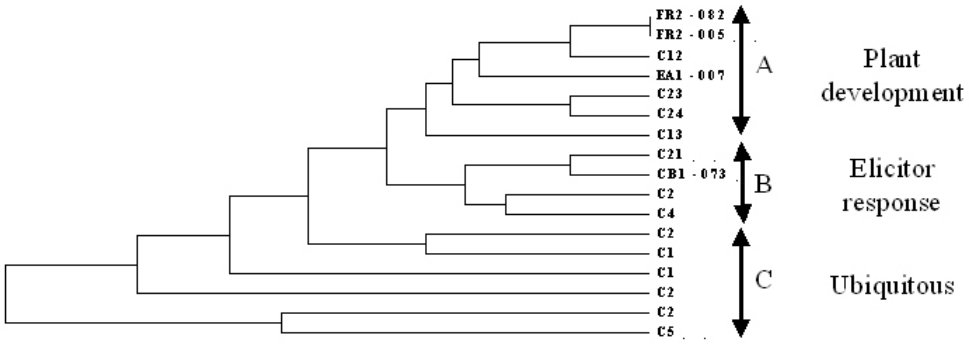

Figure 2. Dendrogram showing relationships among *C. arabica WRKY* unigenes based on their expression data.

REFERENCES

Dong J., Chen C. and Chen Z.. 2003. Plant Mol. Biol. 51:21-37.

Eulgem T., Rushton P.J., Robatzek S. and Somssich I.E. 2000. Trends Plant Sci. **5**:199-206.

Fernandez D., Santos P., Agostini C. et al. 2004. Mol Plant Pathol., 5, 527-536.

Ganesh D., Petitot A.-S., Silva M. et al. 2006. Plant Science, 170:1 045-1051.

Lecouls A.-C., Petitot A.-S. and Fernandez D. 2006. Proc. XXI Scientific Colloquium on Coffee, ASIC, Montpellier.

Lin C., Mueller L.A., Mc Carthy J. et al. 2005. Theor. Appl. Genet. 112:114-30.

Petitot A.-S., Lecouls A.-C. and Fernandez D. 2006. Phylogenetic origins and expression analysis of a duplicated *WRKY* gene in the polyploid species *Coffea arabica.* Proc. XXI Scientific Colloquium on Coffee, ASIC, Montpellier.

Ryu H.S., Han M., Lee S.K. et al. 2006. Plant Cell Rep. DOI10.1007/s00299-006-0138-1.

Ülker B. and Somssich I.E. 2004. Current Opinion in Plant Biology, 7, 491-498.

Vieira L. G. E., Andrade A. C., Colombo C.A. et al. 2006. Brazilian Journal of Plant Physiology, 18, 95-108.

Van de Peer Y. and De Wachter R. 1994. Comput. Applic. Biosciences 10, 569-570.

Xie Z., Zhang Z.L., Zou X. et al. 2005. Plant Physiol, 137:176-189.