Genetic patterns of Vitellaria paradoxa: SNP *vs* Chemical traits in relationship with environment caracteristics

F. Allal, L. Millet, V. Vaillant, F. Davrieux, G. Piombo and J.-M.Bouvet
CIRAD BIOS, UPR39 Research Unit "Genetic Diversity and Breeding of Forest Tree Species"
TA A39/C, 34398 Montpellier Cedex 5, France.
Contact: francois.allal@cirad.fr
jean-marc.bouvet@cirad.fr

Introduction

Chemical components of seeds have an important impact on tree adaptation and future distribution of species in the perspectives of climatic changes. In this study, we try to address the historical, environmental, climatic and genetic causes of the variation of seed's composition in *Vitellaria paradoxa*, a major species in African agroforestry systems. We present here results concerning the great variability observed in chemical trait variation, a phylogeny reconstruction phylogeography of the species and preliminary results concerning the genetic diversity patterns of SAD gene (a key gene responsible to stearic/oleic acid ratio) and VTE2-2 (strongly involved in tocopherol biosynthesis pathway).

Material and Methods

• Material : DNA was extracted from 110 trees belonging to 44 populations distributed on 12 countries. From these 110 genotypes, 24 individuals have been used for gene sequencing (Figure 1).

To determine chemical composition, seeds were collected from 600 individuals following a North to South axis in each country sampled (Senegal, Mali, Bukina, Ghana and Uganda).

• Phenotype determination : After a petrol-ether extraction of seed's fat, Fatty Acid (FA) composition of seeds was determined by CPG analysis, and tocopherol content by HPLC.

• Neutral diversity analysis and Biogeography phylogeny assessment : 8 nuclear SSR markers were used to assess neutral diversity on the natural area. Several markers including 5 cpSSR and 2 chloroplastics sequences were used in order to propose a phylogenetic history address phylogeography of *Vitellaria paradoxa*.

• Candidate genes diversity : SAD gene diversity was assessed by cloning and sequencing Intron 2 of the gene, as Exons 7 to 9 of VTE2-2 (Homogentisate phytyl transferase) was directly sequenced from PCR products.

Figure 1 est ce utile

Di[illegible]

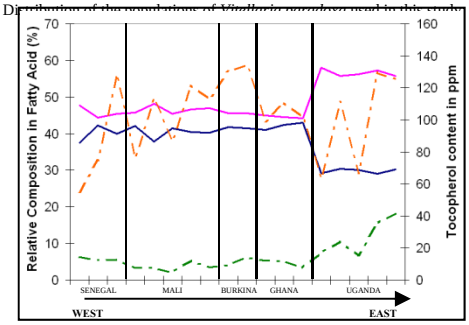

Figure 2 : Relative composition of Oleic (pink) and Stearic (bleu) acids, and Alpha (Orange) and Gamma (Green) tocopherol content (ppm) in seeds.

Results & Discussion

• Chemical Traits Variations : A huge variation in seeds content was has been observed through the natural area (Figure 2). We expected a gradient variation for Oleic/Stearic acids and tocopherol content from North to South in each country as suggested by the results of Marranz *et al* (2003 and 2004), but this reliability is not clear trend was not observed in our study. The variability could be due to local environmental characteristics.

• Neutral diversity and Biogeography of *Vitellaria paradoxa*: The study of 8 nuclear SSR markers revealed a very low F_{st} (0.08) across the natural area, suggesting a huge gene flow and a weak genetic drift. Nucleotide polymorphism of TrnK and TrnQ from chloroplast genome revealed only 3 SNPs distributed on 2kb and did not allowed us to build a clear phylogeny of the species. The 5 cpSSR allowed us to test evolution history of the species as suggested by the chlorotype distribution (figure 3 and 4). This result suggested at the Western part of the area, a large seed flow, and at the Eastern part, a strongly structured genetic diversity. This could result from a combined theory of glacial refuges, human practices in the western part, and geographical barriers at the Eastern part of the area.

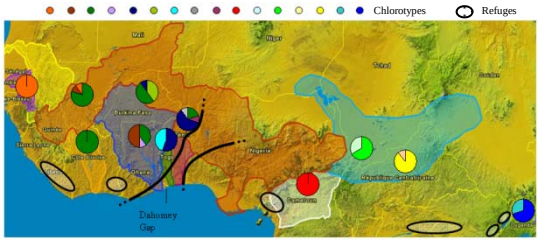

Figure 3 : Chlorotypes distribution and potentials refuges on the natural area

• Description of the nucleotidic variability in the SAD genes : We found 2 expressed SAD genes in *Vitellaria paradoxa*. This duplication has never been observed yet in plants. This discovery has been then a huge problem for genotyping, as the two copies are very closed from each other. Only 8 SNPs and 2 InDels have been revealed between the 2 copies on 2kb comparison (Figure 5). One of the copies (SAD1) seems non-functional, as *in silico* translation lead to a truncated protein, due to a stop codon at the end of the first exon. Based on the polymorphism observed in Intron 2, SAD1 seems also to accumulate twice more mutations than SAD2. VTE2-2 (Figure 6) shows a weak nucleotide diversity, with only 5 SNPs detected on 52 haplotypes. A summary of nucleotid variation is shown Table 1.

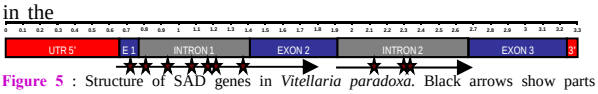

Figure 5 : Structure of SAD genes in *Vitellaria paradoxa*. Black arrows show parts sequenced in our study and red stars positions segregating the to copies SAD1 and SAD2.

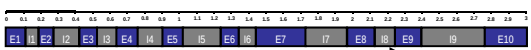

Figure 6 : Gene structure of VTE2-2 in *Vitellaria paradoxa*. Black arrow shows sequenced part in our study

	Region	N	Sites	SNPs	SNPs Frq	Singl.	Singl. Frq	Hap	π	Tajima D	Fu&Li D*	Fu&Li F*
SAD 1	Intron 2	**22**	**779**	**30**	**1/26**	**8**	**1/97**	**16**	**0.00896**	**-0.5837 (ns)**	**0.1205 (ns)**	**-0.1088 (ns)**
SAD 2	Intron 2	**38**	**779**	**32**	**1/24**	**25**	**1/31**	**24**	**0.00273**	**-2.5142 (***)**	**-3.9407 (**)**	**-4.0942 (ns)**
VTE2-2	Exon 7 to Exon 9	**52**	**765**	**5**	**1/153**	**4**	**1/191**	**6**	**0.0003**	**-1.9057 (*)**	**-2.783 (*)**	**-2.937 (*)**

Table 1 : Polymorphism summary and neutrality tests (highlighted in red) for SAD1, SAD2 and VTE2-2 genes.
ns = non-significant ; # = p<0.1 ; * = p<0.05 ; ** = p<0.01 ; *** = p<0.001 (*DNAsp software, Rozas et al. 2003*)

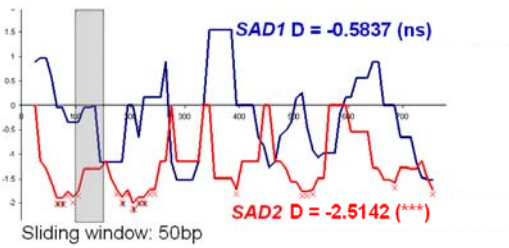

Figure 6 : Sliding Window plot of SAD1 and SAD2 Tajima's D using 50bp in widow length and a step size of 10 bp (*DNAsp software, Rozas et al. 2003*)

• Natural Selection Assessment : We calculated Tajima's D (D) for Intron 2 of the both SAD copies. Although these are non-coding sequences, we can notice that SAD1 exhibit non significant negative Tajima'S D, whereas SAD2 shows a largely significant D=-2.5142 (p<0.001) (Table 1) suggesting this sequence is under a purifying selection. We plotted D using a sliding window of 50bp length and a step size of 10 sites (Figure 6). We can see that high T-D significance is due to the first 200 bp of the Intron 2. The end of Exon 2 contains the most important domains for the protein activity (ferritin-like, FA desaturase). Then we can suppose that due to a high Linkage Disequilibrium, a purifying selection on Exon 2 important domains, could have led to Intron 2 selection. We calculated also D for Exon 7 to 9 of VTE2-2 gene. This gene shows as SAD2 a significant and negative Tajima's D resulting from a purifying selection. However, these results must be considered with caution because negative Tajima's D could also result from a demographic process, as population expansion. These results need to be completed by an analysis of diversity at a smaller scale level (a country), in order to limit demographic effects on the polymorphism observed.

Perspectives

- SNPs typing of the remaining *Vitellaria paradoxa* genotypes and smaller scale gene typing.
- Association studies: Chemical traits/Genotypes for North-South gradients.
- Neutral marker genotyping of new populations from Nigeria and Soudan to better assess phylogeography pattern.

Bibliography

CIRAD BIOS, UPR39 Forest Genetics unit
TA A39/C, 34398 Montpellier Cedex 5, FRANCE.
Web site: www.cirad.fr

IUFRO Tree Biotechnology International Conference Whistler, British Columbia, Canada 2009-06-28 au 2009-07-02