



Identification de marqueur 'InDel' au niveau du gène *Cu-Zn-superoxidedismutase (SOD)* par des approches moléculaires et bioinformatiques

Rayda Ben Ayed^{1*}, Mohamed Ali Triki², Ahmed Rebai¹

¹ Laboratory of Molecular and Cellular Screening Processes, Center of Biotechnology of Sfax, University of Sfax, Sidi Mansour Road, P.O. Box 1177, 3018, Sfax, Tunisia.

² Laboratory of Improvement and Protection of Olive Genetic Resources, University of Sfax, Olive Tree Institute, Aeroport road, 3000, Sfax, Tunisia

* Corresponding Author: raydabenayed@yahoo.fr

Received 29 December 2021; Revised 13 April 2022; Accepted 20 April 2022

Abstract

The *Cu-Zn-superoxide dismutase (SOD)* gene encodes an enzyme involved in the detoxification of superoxide radicals. The genotyping of this gene in 16 Tunisian olive varieties allowed the identification of an InDel-type SNP marker but also to validate the genotype-phenotype relationship between this marker and some quantitative (yield) and qualitative (olive oil quality) parameters. The phylogenetic analysis revealed a clustering into two groups independently of geographical origin. On the basis of bioinformatics analyses, highly significant associations were observed for varieties with the ID genotype producing more linoleic acid and for varieties with the II genotype having a high cholesterol level. This may suggest that varieties with genotype II (mainly Oueslati and Chemlali Tataouine) produce more oleic acid than linoleic acid than those with homozygous DD and heterozygous ID genotypes.

Keywords: *Olea europaea* L., Cu-Zn-superoxide dismutase (SOD), genotyping, InDel, logistic regression.

1. Introduction

L'olivier (*Olea europaea* L.) est un arbre rustique, typiquement méditerranéen, cultivé depuis l'Antiquité pour ses fruits et l'huile d'olive qu'il produit. Cependant, bien que la plupart des variétés d'olivier résistent bien à la sécheresse estivale, au froid et à la salinité, ce n'est pas le cas pour certaines variétés (Ben Ayed *et al.*, 2015).

Les contraintes environnementales, y compris le stress hydrique, engendrent l'apparition d'un stress oxydatif, dû à l'accumulation d'espèces réactives de l'oxygène (ROS), qui endommagent les structures cellulaires (Appel et Hirt, 2004). Dans des conditions normales, les feuilles sont dotées d'enzymes et de métabolites antioxydants suffisants pour faire face aux ROS. Ces enzymes telles que les superoxidedismutases (SOD) sont aussi exprimées pendant le stress hydrique (Flexas

et al., 2006). Les Cu-Zn Superoxidedismutases sont des métalloprotéines ayant une activité enzymatique qui catalysent la dismutation du superoxyde en oxygène et peroxyde d'hydrogène. Cette enzyme est une partie importante du système de défense contre les radicaux libres. Elle assure aussi la protection des tissus biologiques de tous les ions superoxide (Alscher *et al.*, 2002). En effet, Les plantes se protègent par l'utilisation de toute une série d'enzymes et molécules antioxydantes pour détoxifier les espèces réactives de l'oxygène produites (Alscher *et al.*, 2002).



L'objectif de ce travail est d'étudier un marqueur SNP de type InDel localisé au niveau du gène *Cu-Zn-superoxidedismutase (SOD)* afin de classier 16 variétés d'olivier tunisiennes d'une part et de valider la relation génotype-phénotype entre ce marqueur et quelques paramètres quantitatifs et qualitatifs d'autre part, pouvant aboutir à l'identification d'éventuelles associations entre variabilité moléculaire et paramètres de qualité des huiles et d'investiguer l'influence de la présence de l'olivier dans des conditions de stress sur la qualité de l'huile d'olive.

2. Matériel et méthodes

2.1. Matériel biologique

Seize variétés tunisiennes d'olivier ont été sélectionnées dans différentes régions géographiques du pays allant du nord au sud et couvrant ainsi plusieurs étages bioclimatiques (Tableau 1).

Tableau 1. Liste des seize variétés étudiées et leur provenance.

Variétés	Abréviation	Régions
Chemlali_Chàal	Chem_Chàal	Sfax
Chemlali_blettech	Chem_Blett	Sfax
Chemlali_Sidi-bouزيد	Chem_SB	Sidi Bouzid
Chemlali_Sousse	Chem_Sous	Sousse
Chemlali_Dokhane	Chem_Dok	Sfax
Chemlali_Monastir	Chem_Mon	Monastir
Chemlali_Nabeul	Chem_Nab	Nabeul
Chemlali_Tataouine	Chem_Tat	Tataouine
Chetoui_Seliana	Chet_Silia	Siliana
Chetoui_Thibar	Chet_Thibar	Beja
Chemcheli	Chemch	Gafsa
Oueslati	Oueslati	Kairouane
Zarrazi	Zarrazi	Zarzis
Zalmati	Zalmati	Zarzis
Rkhaymi	Rkhaymi	Nabeul
Chetoui_Nabeul	Chet_Nab	Nabeul

2.2. Extraction de l'ADN génomique de l'olivier

De jeunes feuilles fraîchement prélevées sont utilisées pour extraire l'ADNg en utilisant le protocole CTAB comme décrit par Fabbri *et al.* (1995) et auquel nous avons apporté une légère modification. L'ADN génomique (non dilué) a été stocké dans un tampon TE (1X pH 8,0) à -20 °C.

2.3. Conception *in silico* des amorces de la PCR

Nous avons choisi l'ADNc codant pour l'enzyme Cu-Zn-superoxydismutase (numéro d'accèsion : AF426829) à partir de la base de donnée publique GeneBank du site NCBI (<http://www.ncbi.nlm.nih.gov/>). Nous avons désigné des amorces pour cet ADNc à l'aide du programme Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) afin d'amplifier un fragment de l'ADNc choisi et correspondant au site du SNP. Après une étape d'optimisation, nous avons réalisé une amplification par PCR des séquences en utilisant les amorces nouvellement conçues dans cette étude.

2.4. Amplification par PCR

La validation du couple d'amorces PCR a consisté à l'amplification des fragments contenant le marqueur SNP d'intérêt au moyen de la réaction de PCR.

La réaction de PCR est réalisée dans un volume réactionnel de 30µl dans un thermocycleur de type GeneAmp PCR System 9700 (PERKIN-ELMER), contenant : 10µl Tampon Promega (5X), 2,5µl MgCl₂ (25mM), 2µl dNTP (10mM), 1µl d'oligonucléotide sens (10µM), 1µl d'oligonucléotide anti-sens (10µM), 10 µl d'ADN (10ng/µl), 0,25µl Taq polymérase (5U/µl) et 3,25µl H₂O.

Les conditions du programme de la réaction PCR sont: une dénaturation initiale 94°C pendant 1 minute, puis 35 cycles réactionnels (Dénaturation : 94°C pendant 40 secondes, Hybridation : 57°C pendant 30 secondes, Elongation : 72°C pendant 45 secondes) et une élongation finale 72°C pendant 10 minutes. La taille des fragments amplifiés est de 190pb et 176pb pour insertion et délétion, respectivement. Une simple PCR permet de détecter l'insertion ou la délétion.

2.5. Analyses statistiques et bioinformatiques

L'étude phylogénétique a été effectuée en utilisant le programme NTSYS-pc version 2.1 et la méthode UPGMA (Unweighted Pair Group Method using an Arithmetic average). L'étude d'association génotype-phénotype a été effectuée moyennant le langage R. Cette étude est utilisée pour évaluer les relations entre paires de variables notées selon une échelle binaire en utilisant l'analyse de variance. Les associations ayant une valeur P < 0.05 sont considérées significatives. Un test de Student a été également utilisé pour évaluer des associations significatives avec les variables quantitatives. Une analyse d'apprentissage automatique (Machine learning) a été également réalisée en utilisant la régression logistique pour évaluer les relations entre une variable d'intérêt notée (le marqueur SOD) et plusieurs autres variables.

3. Résultats et discussion

3.1. Typage du marqueur InDel localisé dans le gène *SOD*

3.1.1. Préparation *in silico* des amorces et amplification PCR

Nous avons choisi des ADNc car le génome de l'olivier n'est pas complètement séquencé et par conséquent les séquences génomiques complètes de ce gène ne

sont pas disponibles dans le site NCBI. Nous avons désigné des amorces pour cet ADNc à l'aide du programme Primer 3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) afin d'en amplifier un fragment précis. Un marqueur InDel (polymorphisme de type insertion/délétion) localisé au niveau du gène *Cu-Zn superoxidedismutase (SOD)* a été choisi pour être étudié. Après une étape d'optimisation, nous avons réalisé une amplification par PCR des séquences en utilisant les amorces nouvellement conçues *in silico* (Figure 1). La PCR permet de détecter l'insertion ou la délétion.

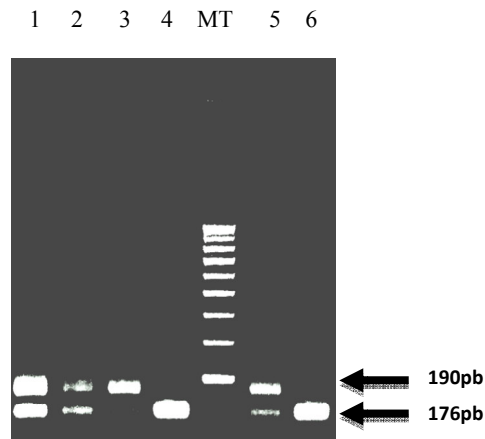


Figure 1. Amplification par PCR des séquences d'ADN spécifiques du SNP de type insertion/délétion (SOD) sur gel d'agarose 2% - 1: *Chemlali_Sousse*; 2: *Zalmati*; 3: *Oueslati*; 4: *Koroneiki*; 5: *Chetoui_Nabeul*; 6: *Picholine*; MT: Marqueur de taille (ADN 100 pb StepLadder, Promega)

3.1.2. Caractéristiques du marqueur InDel étudié

Après le typage du marqueur SNP de type insertion/délétion, nous avons déterminé avec précision le génotype de chaque variété étudiée. Nous avons obtenu 3 génotypes, II : insertion ; DD : délétion et ID : insertion/délétion.

Les valeurs des fréquences génotypiques ainsi que celles des fréquences alléliques du marqueur étudié sont mentionnées dans les figures 2 A et 2B, respectivement.

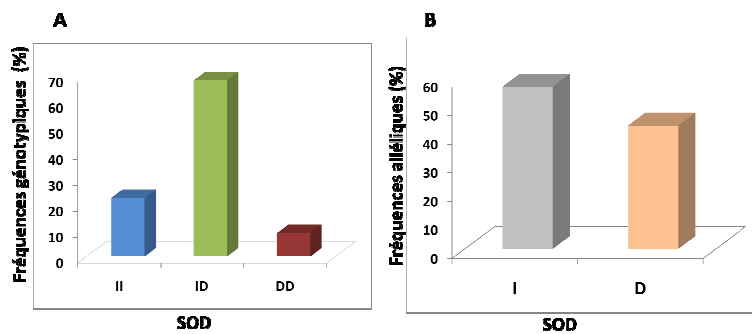


Figure 2. Fréquences génotypiques et alléliques du marqueur SOD étudié

Nous pouvons déduire d'après le résultat de la figure 2B que l'allèle I de polymorphisme de type insertion/délétion paraît prédominant dans notre échantillon de variétés d'olivier, ce qui suppose qu'il pourrait s'agir d'allèle ancestral. Le calcul des fréquences génotypiques (Figure 2A) montre la présence des trois génotypes pour le marqueur SOD.

L'hétérozygotie observée pour le marqueur SOD est de 0,681 tandis que l'hétérozygotie théorique est de 0,490. On remarque que l'hétérozygotie observée est supérieure à celle attendue (Tableau 2). Le pouvoir de discrimination (PD) est de 0,475. Des valeurs similaires de pouvoir de discrimination ont été trouvées par Real *et al.* (2006). Pour les fréquences alléliques (Figure 2B) du marqueur SOD étudié, nous avons remarqué qu'il y a une dominance d'un allèle I par rapport à l'autre. Nous avons remarqué que la majorité des variétés étudiées ont un génotype hétérozygote avec une fréquence des génotypes hétérozygotes de 68,18% et celle des génotypes homozygotes II et DD de 22,72% et 9,09% respectivement (Figure 2A).

Tableau 2. Caractéristiques du marqueur SOD chez 16 variétés d'olivier^α.

Locus	Type de polymorphisme	Tm ^a	Longueurs des bandes en pb	Ho	He	PD	r
SOD	Insertion/délétion	57	176/128	0,681	0,490	0,475	-0,128

^α: pour chaque locus les tailles des allèles en paires de base, le nombre d'allèles (Na), l'hétérozygotie observée (Ho), l'hétérozygotie attendue (He), le pouvoir de discrimination (PD), et la probabilité d'allèles nuls (r) sont signalés. ^a: température d'hybridation pour chaque PCR.

3.2. Etude phylogénétique des variétés d'oliviers étudiées

Les données génotypiques marqueur InDel ont été utilisées pour construire le dendrogramme représenté dans la figure 3.

Deux groupes de variétés peuvent être définis.

- ✓ Le groupe 1 comprend 3 variétés : Chem_Tat, Chemch et Oueslati.
- ✓ Le deuxième groupe contient les 13 autres variétés : Chem_Chaàl, Chem_Blett, Chem_Dok, Chem_SB, Chem_Nab, Chem_Sous, Chem_Mon, Chet_Thibar, Chet_Nab, Chetoui_Silia, Rkaymi, Zarrazi et Zalmati.

En examinant le dendrogramme, on remarque l'absence d'une structuration de la variabilité génétique selon l'origine géographique.

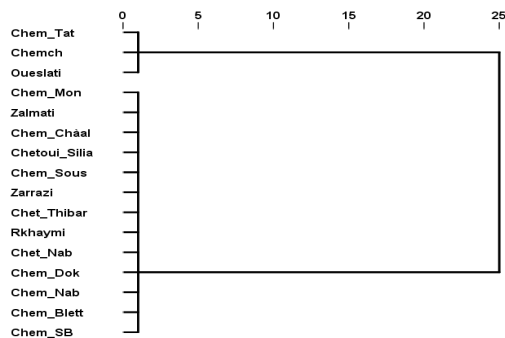


Figure 3. Dendrogramme des génotypes des 16 variétés tunisiennes étudiées par le marqueur InDel en utilisant le programme NTSYS-pc version 2.1 et l'analyse par la méthode UPGMA.

3.3. Etude d'association entre les variants du marqueur InDel du gène *SOD* et les paramètres de qualité de l'huile d'olive

Le marqueur polymorphe *SOD* est de type insertion/délétion, il est localisé dans le gène *Cu-Zn-superoxidedismutase* impliqué dans le mécanisme de défense de la plante contre le stress oxydatif (Alscher *et al.*, 2002) et hydrique (Flexas *et al.*, 2006). Dans cette partie, nous nous sommes intéressés à évaluer les relations entre les différents paramètres de l'huile d'olive et le marqueur InDel étudié en utilisant l'analyse de variance (après correction de Bonferroni pour tests multiples).

L'analyse d'association génotypique de ce marqueur avec l'ensemble des paramètres quantitatifs n'a montré aucune association significative, à l'exception de la teneur en acides gras polyinsaturés type C18:2. Cette relation est hautement significative dans les variétés qui ont le génotype ID produisant plus d'acide linoléique ($p=0,008$) (Tableau 3b). Egalement, nous avons observé une relation significative avec les variétés avec le génotype II avec un taux élevé en Cholestérol ($p=0,045$) (Tableau 3a).

Tableau 3 (a,b). Association entre les génotypes *SOD* et les paramètres quantitatifs de l'huile d'olive

(a) Paramètre	SNP	Moy ±DS	P
Cholestérol	ID	0.0375±0.034	0.045
	II	0.083±0.011	
β-sitostérol	ID	84.10±0.649	0.9
	II	84.03±1.563	
Chlorophylles	ID	2.313±1.083	0.519
	II	2.767±0.850	
Carotènes	ID	5.239±1.537	0.598
	II	4.733±0.907	
PMF	ID	1.166±0.831	0.933
	II	1.616±0.752	
(b) Paramètre	SNP	Moy ±DS	P
Acide oléique C18 :1	ID	0,226±0,031	0,230
	II	0,250±0,10	
Acide linoléique C18 :2	ID	16,31±2,938	0,008
	II	10,87±0,961	
Acide linoléique C18 :3	ID	0,638±0,619	0,312
	II	0,600±0. 0	
Taux d'insaponifiable	ID	1,259±0,095	0,235
	II	1,027±0,586	
Stabilité oxydative	ID	45,24±15,250	0,191
	II	57,73±4,244	

3.4. Apprentissage automatique par régression logistique

L'association du marqueur InDel avec la teneur en C18:2 ($p=0,007$) est confirmée par la méthode d'apprentissage automatique (Machine learning) en utilisant l'algorithme de régression logistique (Tableau 4). Ces résultats sont biologiquement appropriés vu que ce marqueur est localisé dans le gène SOD qui est impliqué dans la défense de l'arbre contre les conditions stressantes et par la suite pourra influencer la qualité de l'huile d'olive notamment la composition en acides gras.

Tableau 4. Résultats d'associations génotype-phénotype en utilisant l'apprentissage automatique par régression logistique.

Modèle	SOD
PMF	0,981
Acidité	0,289
Teneur en polyphénols totaux (mg/kg)	0,450
Taux d'insaponifiable (%)	0,069
Stabilité oxydative (h à 100°C)	0,273
Teneur en chlorophylle (ppm)	0,676
C18 :2	0,007
C18 :3	0,248
Cholestérol	0,103
Teneur en carotènes (ppm)	0,607
β -sitostérol	0,090
Maturation de fruit	0,861
Utilisation	0,649
Forme de fruit	0,672
Pollinisation	0,672
Tolérance de l'arbre	0,425
Teneur en huile par fruit	0,243
Productivité de l'arbre	0,274

*: on donne les p-valeurs dans l'analyse de régression logistique

4. Conclusion

L'étude du marqueur SNP de type InDel localisé au niveau du gène *Cu-Zn-superoxidedismutase* permet de valider la relation génotype-phénotype entre ce marqueur et quelques paramètres quantitatifs et qualitatifs de l'huile d'olive, pouvant ainsi d'aboutir à l'identification d'éventuelles associations entre variabilité moléculaire et paramètres de qualité des huiles d'une part et de l'adaptation de l'arbre d'autre part. En effectuant des analyses bioinformatiques et statistiques, quelques associations très significatives ont été observées pour les variétés qui ont le génotype ID produisant plus d'acide linoléique et pour les variétés ayant le génotype II possédant un taux élevé en cholestérol. L'apprentissage automatique par un algorithme de régression logistique confirme l'association avec l'acide linoléique. Ceci pourra suggérer que les variétés qui possèdent le génotype II (principalement Oueslati et Chemlali Tataouine) produisent plus d'acide oléique que d'acide

linoléique que celles qui ont des génotypes homozygotes DD et hétérozygotes ID. Ces résultats d'associations génotype/phénotype pourraient avoir des retombées intéressantes et valorisables sur les plans agronomiques et économiques, notamment pour des études de sélection et d'amélioration variétales, ainsi que l'obtention de génotypes d'olivier résistants aux contraintes abiotiques et biotiques.

Références

Alscher, R.G., Erturk, N., Heath, L.S. 2002. Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *J Exp Bot* 53(372): 1331-41.

Apel, K., Hirt, H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annual Review of Plant Biology* 55: 373-399.

Ben Ayed, R., Grati, N., Moreau, F. Rebai, A. 2009. Comparative study of microsatellite profiles of DNA from oil and leaves of two Tunisian olive cultivars. *European Food Research and Technology* 229 (5): 757-762.

Ben Ayed, R., Ennouri, K., Ben Hassen, H., Triki M.A., Rebai A. 2015. Comparaison between DNA-based, pomological and chemical markers accomplished by bioinformatic tools to distinguish within Tunisian olive cultivars. *J Fundam Appl Sci.* 7(3): 408–421.

Fabbri, A., Hormaza, J.I., Polito, V.S. 1995. Random Amplified Polymorphic DNA Analysis of Olive (*Olea europaea* L.) Cultivars. *J. Amer. Soc. Hort. Sci.* 120(3): 538-542.

Flexas, J., Bota, J., Galmés, J., Medrano, H. and Ribas-Carbó, M. 2006. Keeping a positive carbon balance under adverse conditions: responses of photosynthesis and respiration to water stress. *Physiologia Plantarum* 127:343-352.

Citation: Ben Ayed, R., Triki, M.A, Rebai, A. 2022. Identification de marqueur 'InDel' au niveau du gène *Cu-Zn-superoxydodismutase (SOD)* par des approches moléculaires et bioinformatiques. *J.A.A.O.G* 1(2):1-8.