In Silico Identification and Characterization of Fatty Acid Desaturase (FAD) Genes in Argania spinosa L. Skeels: Implications for Oil Quality and Abiotic Stress

Abdelmoiz El Fager, Karim Rabeh, Mohammed Alami, Abdelkarim Filali-Maltouf and Bouchra Belkadi

Team of Microbiology and Molecular Biology, Plant and Microbial Biotechnology, Biodiversity and Environment Research Center, Faculty of Sciences, Mohammed V University, Rabat, Morocco.

Bioinformatics and Biology Insights Volume 18: 1-15 © The Author(s) 2024 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/11779322241248908



ABSTRACT: Fatty acid desaturase (FAD) is the key enzyme that leads to the formation of unsaturated fatty acids by introducing double bonds into hydrocarbon chains, and it plays a critical role in plant lipid metabolism. However, no data are available on enzyme-associated genes in argan trees. In addition, a candidate gene approach was adopted to identify and characterize the gene sequences of interest that are potentially involved in oil quality and abiotic stress. Based on phylogenetic analyses, 18 putative FAD genes of Argania spinosa L. (AsFAD) were identified and assigned to three subfamilies: stearoyI-ACP desaturase (SAD), Δ-12 desaturase (FAD2/FAD6), and Δ-15 desaturase (FAD3/FAD7). Furthermore, gene structure and motif analyses revealed a conserved exon-intron organization among FAD members belonging to the various oil crops studied, and they exhibited conserved motifs within each subfamily. In addition, the gene structure shows a wide variation in intron numbers, ranging from 0 to 8, with two highly conserved intron phases (0 and 1). The AsFAD and AsSAD subfamilies consist of three (H(X)2-4H, H(X)2-3HH, and H/Q (X)2-3HH) and two (EEN(K)RHG and DEKRHE) conserved histidine boxes, respectively. A set of primer pairs were designed for each FAD gene, and tested on DNA extracted from argan leaves, in which all amplicons of the expected size were produced. These findings of candidate genes in A spinosa L. will provide valuable knowledge that further enhances our understanding of the potential roles of FAD genes in the quality of oil and abiotic stress in the argan tree.

KEYWORDS: Fatty acid desaturase genes, candidate gene, quality oil, abiotic stress, A spinosa L

RECEIVED: September 26, 2023. ACCEPTED: April 4, 2024.

TYPE: Original Research Article

FUNDING: The author(s) received no financial support for the research, authorship, and/or publication of this article

Introduction

The argan tree (Argania spinosa L. Skeels) is a species endemic to Morocco in the Sapotaceae family. These species are known for their resistance to drought and heat because of their abundance in arid and semi-arid areas of southwest Morocco.¹ It is a valuable resource for a country's economy and environment.

Argan oil, now widely recognized as one of the most expensive woody oil plants, has traditionally been used in local cuisine, medicine, and cosmetics. It has also been used to prevent certain cardiovascular diseases. Its advantages are due to its balanced composition of fatty acids, principally oleic, linoleic, and palmitic acids, and other minor compounds, including polyphenols, tocopherols, and sterols, which confer antioxidant properties.² However, linoleic and α -linolenic acids cannot be synthesized by the human body and must be obtained from dietary sources.3

With their function as energy reserves and status as the primary components of plant cells, fatty acids serve numerous roles in plant metabolism, membrane function, hormonal signaling, and plant development.⁴ The fatty acid composition of argan oil shows a predominance of oleic and linoleic acids.⁵ According to its composition, argan oil is an oil of the oleic/ linoleic type consisting of almost 80% unsaturated fatty acid. DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Bouchra Belkadi, Team of Microbiology and Molecular Biology, Plant and Microbial Biotechnology, Biodiversity and Environment Research Center, Faculty of Sciences, Mohammed V University, Rabat 10106, Morocco. Email: b.belkadi@um5r.ac.ma

Palmitic acid constitutes the largest proportion of saturated fatty acids, whereas linolenic acid constitutes less than 0.25%.⁶

Oil quality depends on the content of unsaturated fatty acids.⁷ The biosynthesis of fatty acids represents the key step in the formation of the oil body, so the enzymes involved in elongation and desaturation play a major role in the formation of fat, which are regulated by several genes that give them a predisposition in terms of quality.8 These genes belong to the family of fatty acid desaturases (FADs).9 The desaturation reactions carried out by FADs are essential for the production of unsaturated fatty acids, because they lead to the formation of double bonds.¹⁰ Based on their solubility, FADs are classified into two main groups: membrane-bound desaturases and soluble desaturases.¹¹ Stearoyl-ACP-desaturase (SAD) is the only soluble FAD desaturase found in the plastid matrix, whereas membrane-bound FAD desaturases are present in a range of organisms including plants, algae, animals, and fungi.12 SAD, FAD4, FAD5, FAD6, FAD7, and FAD8 are groups of enzymes found in plastids that play vital roles in lipid desaturation reactions.¹³ In contrast, FAD2 and FAD3 are located in the endoplasmic reticulum (ER) and are responsible for directing the desaturation of lipids that are not located in the chloroplast.14 The conversion of oleic acid (18:1) to linoleic acid (18:2) is catalyzed by FAD2 in the ER and

 $(\mathbf{\hat{n}})$

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (https://creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage). *FAD6* in the plastid, while the desaturation of linoleic acid (18:2) to linolenic acid (C18:3, n6) takes place in the ER and plastids by *FAD3* and *FAD7/FAD8*, respectively.⁷ *FAB2/SAD* introduces a double bond to saturated stearic acid (18:0) to convert it into unsaturated oleic acid (18:1).¹⁵ There are two electron donors that lead to the formation of the double bond: NADPH/ferredoxin, and NADH/cytochrome b5 found in the chloroplast and ER, respectively.¹⁶

The plant *FAD*s have equal significant functions in responding to abiotic stresses, including high temperatures, salt, and cold.¹⁷ Specifically, studies have demonstrated that the overexpression of *FAD2* and *FAD6* in Arabidopsis significantly increased tolerance to salt stress.^{18,19} Furthermore, overexpression of *FAD7* contributes to the reduction of damage caused by cold stress in tobacco.²⁰ Similarly, *NtFAD3* and *NtFAD8* can promote tobacco drought stress.²¹ To date, numerous genome sequencing studies conducted on various plants have served as a crucial resource for the genome-wide analysis of the *FAD* gene family in these plant species. For instance, studies have identified 27 *FAD* genes in wild olives,²² 23 in cucumbers,²³ 27 in bananas,⁴ and 21 in chickpeas.²⁴ However, no information is available on this gene family in *A spinosa* L.

To gain a deeper understanding of their evolution and to improve their role in tolerance to abiotic stress, the identification and characterization of FAD genes from the published Argan genome are necessary. Therefore, in this study, we present the first genome-wide identification of FAD genes in A*spinosa* L. To this end, we investigated the in silico identification of FAD genes and subsequently characterized their phylogenetic relationships, gene structure, transmembrane (TM) domains, conserved motifs, and three-dimensional (3D) modeling. In addition, we developed specific primers from the identified Argan FAD genes and validated numerous DNA accessions of A spinosa L. species.

Materials and Methods

Identification of Argan FAD gene family and sequences analysis

To identify all members of the *FAD* and *SAD* genes in *A* spinosa L., 33 sequences encoding *FAD* and *SAD* proteins from oil plants, such as *O europaea* L., *Sesamum indium* L., and *Arabidopsis thaliana* L., were retrieved from the NCBI database (Supplemental Table S1). These sequences were used as queries to confront them with the Argan genome using the BLAST tool (tBLASTn) with the *E*-value set to 1×10^{-5} . The contigs obtained from the BLAST results were analyzed using the AUGUSTUS tool to obtain the coding and complete sequences of the predicted *FAD* and *SAD* genes. All candidate sequences were further analyzed using the Pfam tool (http://pfam.xfam.org) to confirm all putative *AsFAD* members by the presence of the domain kept under the name "desaturase". The physicochemical features of the *AsFAD* and *AsSAD* proteins were detected using the

ProtParam tool ExPASy (http://web.expasy.org/protparam/).²⁵ The subcellular localization of the proteins was analyzed using WoLFPSORT (https://wolfpsort.hgc.jp/) and Plant-mPLoc (http://www.csbio.sjtu.edu.cn/bioinf/ plant-multi/).^{26,27}

Gene structure and protein motif analysis of AsFADs family

Intron-exon structures and the type of splicing of genes in the *AsFAD* family were visualized using the Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/) by comparing the complementary DNA sequences with their corresponding genomic DNA sequences retrieved from the AUGUSTUS gene annotation file.²⁸ The conserved motifs of all putative *AsFAD* proteins were predicted using MEME version 5.4.1 (https://meme-suite.org/tools/meme/), with the following parameters: maximum number of motifs of 6 and 10 for *SAD* and *FAD*s, respectively; minimum motif width ranging from 5 to 50 residues; minimum number of sites as 2; and distribution of motifs as 0 or 1.²⁹ The predicted functionality of the identified motifs was determined using the Pfam server (http:// pfam.xfam.org).³⁰

Multiple sequence alignment and phylogenetic analysis

Multi-alignment of the identified *FADs* from *A spinosa* L. with *FADs* from *O europaea* L., *S indium* L., *Theobroma cacao* L., *Linum usitatissimum* L., and *A thaliana* L. was performed using Clustal Omega. Additional analysis was conducted on the alignment files to identify any conserved histidine-rich boxes. A neighbor-joining phylogenetic tree was constructed using MEGAX.³¹ Bootstrapping was carried out with 1000 replicates for statistical reliability.³²

The prediction of cis-elements within the promoter regions of AsFAD genes

Using TBtools, the 1500 bp regions located before the ATG start codon of each *AsFAD* gene were retrieved from the Argan genome draft database, and subsequently evaluated for potential cis-acting elements and transcription factor-binding sites through the use of PlantCARE (http://bioinformatics.psb. ugent.be/webtools/plantcare/html/).³³

Predicted 3D structures of AsFAD proteins

The 3D structure model and the *AsFAD* template were performed by the Phyre2 Software server (http://www.sbg.bio. ic.ac.uk/~phyre2/html/page.cgi?id=index). The detection rate in the Phyre2 database was used to identify a homology model using hidden markov model (HMM) scanning.³⁴ Templates with the highest-scoring crystal structure were selected for model building. The TM helices were predicted from the topology sequence analyzed using Phyre2.

Designing and validation of primers

The Primer3 online tool (https://bioinfo.ut.ee/primer3/) was used to design primers. All predicted argan gene sequences were used to design the primers flanking the desired parts according to the following thresholds: oligonucleotide length between 16 and 26 nucleotides, GC rate 40% to 70% (ideally around 55%), melting temperature (~60 °C), mononucleotide tracks shorter than 3, product length between 120 and 800 bp, and self-dimerization of primers—not acceptable. In addition, hairpin structure formation and cross-dimerization were verified using the PCR Primer Stats tools (https://www.bioinformatics.org/sms2/pcr_primer_stats.html).

The designed primer pairs were commercially synthesized and validated to test their effectiveness. To accomplish this, genomic DNA was extracted from young leaves of *A spinosa* L. using ISOLATE PLANT KIT (Bioline, London) following the manufacturer's protocol. Subsequently, polymerase chain reaction (PCR) amplification was performed in a 25-µl reaction volume using MyTaq TM HS Mix (Bioline, London) according to the manufacturer's guidelines. The PCR products were visualized on 2% agarose gel.

Results

Identification of FAD genes in A spinosa L

We identified 18 distinct AsFAD genes (non-redundant sequences) in the A spinosa L. genome (9 FADs and 9 SADs). They were named AsFAD and AsSAD according to their phylogenetic analysis results (Supplemental Table S1). Genomic DNA, CDS, and amino acid sequences are shown in Supplemental Figure S1. The full-length DNA sequences of the putative AsFAD genes ranged from 1398 (AsFAD2-a) to 8174bp (AsFAD6). The shortest CoDing Sequence (CDS) length was 999 bp (AsSAD1-4), and the longest one contained 1371bp (AsFAD7-b). The AsFAD and AsSAD peptide sequences contain 332 (AsSAD4) to 456 (AsFAD7-b) amino acids, with a predicted molecular weight of 38.06 kDa to 52.03 kDa and theoretical isoelectric points of 5.19 to 9.32 (Table 1). The cellular localization of proteins was investigated, and the findings indicated that AsSAD demonstrated activity in the ER, whereas AsFAD was found in the chloroplast, except for AsFAD6, which was located in the ER.

Phylogenetic analysis of AsFAD proteins

A neighbor-joining phylogenetic tree was constructed using *FADs* from various oilseed species, such as Olive, Sesame, Cacao, Flax, and a model plant Arabidopsis, to determine the evolutionary relationship among *AsFADs*. The results of the phylogenetic tree construction based on multiple sequence

alignments indicated that the *AsFAD* gene has high homology with other oilseed genes.

Based on the phylogenetic tree, the *AsFAD* proteins were categorized into three separate subfamilies according to their desaturation group (Figure 1). The Δ -15 desaturase group consisted of 14 members, including the *FAD3* and *FAD7* families, in which *A spinosa* L. comprised four members. Both *FAD7* and *FAD3* members were in the same group because of the high homology of their sequences. The Δ -12 desaturase group consisted only of *FAD2* and *FAD6* proteins, and was observed in all oilseed plants included in this study. Furthermore, all the *FAD5* members were present in the front-end desaturase group. However, members of the *SAD* protein group were primarily found in the *SAD* group.

Analysis of gene structure and conserved motifs of AsFAD proteins

To gain further insight into the structural diversity of FADs, we investigated the structures of AsFAD genes. Figure 2 shows the exon-intron organization of the AsFAD and AsSAD genes, and varying positions and lengths were observed among all the genes. The number of exons ranged from 1 to 9. The AsFAD3 and AsFAD6 subfamilies were characterized by nine exons, whereas the AsFAD7 subfamily contained eight exons. However, AsFAD2 contains one and two exons. The number of exons in the AsSAD subfamily ranged from two to four. Most of the AsSAD subfamily contains three exons, with the exception of AsSAD1/4/8, AsSAD6, and AsSAD2 with two, four, and five exons, respectively. All splicing phases (phases 0, 1, and 2) were observed for the AsFAD gene family; AsFAD3, AsFAD6, and AsSAD2 showed all three intron splicing phases, whereas AsFAD7 and AsSAD9 subfamilies had phases 0 and 2. Similarly, AsSAD3, AsSAD5, AsSAD6, and AsSAD7 demonstrated phases 0 and 1.

Using the MEME web server, we identified 10 and 6 conserved motifs in the *AsFAD* and *AsSAD* subfamilies, respectively (Figures 3 and 4). Subsequent analysis with Pfam revealed that both conserved motifs 1 and 2 of *AsFAD* belonged to the "FA_ desaturase," and the function of other motifs was not detected (Supplemental Table S4). Moreover, all six conserved motifs in the *AsSAD* subfamily are associated with the "FA_desaturase_2" (Figure 3). The *AsFAD2* subfamily is composed of eight motifs 1, 2, 3, 4, 5, 6, and 10, while motifs 1 to 9 were harbored in *AsFAD3* and *AsFAD7*. Both the *AsFAD6* and *AsFAD8* subfamily members had only two motifs, 1 and 2. In addition, motif 10 characterized *AsFAD2* members, whereas motifs 8 and 9 were found just in *AsFAD3/7* members (Figure 4).

Based on the multiple sequence alignment (Figures 5 and 6) of the Arabidopsis homologs and the *AsFAD* proteins, three conserved histidine-rich boxes of membrane-bound *FADs* H(X)2-4H (His-Box1), H(X)2-3HH (His-Box2), and H/Q (X)2-3HH (His-Box3) were observed, except for *AsFAD6*,

DICTED	CONTIGS	AMINO	CODING	LENGTH	START	END	INTRON	EXON	₫	MW (KDA)	SUBCELLULAR
		ACID LENGTHS (AA)	SEQUENCE (CDS) LENGTH (BP)	OF GENOMIC DNA (BP)	POSITIONS (BP)	POSITIONS (BP)	NUMBERS	NUMBERS			LOCALIZATIONS
QLO	DD01000980.1	383	1152	1398	6900	8298	0	÷	7.34	44.3	ER
αro	0D01014567.1	387	1164	1795	42836	44631	0	-	8.72	44.91	ER
arc	D01063726.1	384	1155	6497	3826	10323	0	-	8.57	44.09	ER
aro)D01071667.1	434	1305	6098	33786	39884	8	6	8.49	49.52	ER
ard	0D01024743.1	412	1239	5464	168226	173690	8	6	8.79	47.72	ER
QLQ	DD01003815.1	376	1131	8174	91847	100021	8	6	9.32	42.96	Chlo
aro	0D01071352.1	450	1353	3005	65746	68751	7	8	8.45	51.2	ER
ara	DD01003562.1	456	1371	2961	84316	87277	7	8	8.57	52.03	EB
aL	OD01063727.1	440	1323	2645	5836	8481	-	N	8.87	50.24	EB
ara	0D01074982.1	332	666	2228	1543	3771	F	N	5.32	38.14	Chlo
aro)D01074982.1	449	1350	4600	13591	18191	4	ß	7.68	50.49	Chlo
σΓ	OD01074982.1	405	1218	4123	24498	28621	N	e	7.66	46.24	Chlo
g	OD01075084.1	332	666	3504	80117	83621	F	N	5.19	38.06	Chlo
gL	OD01075084.1	390	1173	4539	100972	105511	5	e	6.24	44.6	Chlo
б	OD01062571.1	427	1284	3193	22728	25921	e	4	5.71	48.34	Chlo
σΓ	OD01062447.1	390	1173	4114	1927	6041	N	e	7.66	44.65	Chlo
Ъ	OD01022196.1	387	1164	2112	3232	5344	-	N	6.15	43.61	Chlo
QLQ	DD01071541.1	399	1200	6035	936	6971	N	e	6.68	45.74	Chlo

Abbreviations: Chlo, chloroplast; ER, endoplasmic reticulum; MW, molecular weight; pl, theoretical isoelectric point.

Table 1. Information of the predicted FAD family in Argania spinosa L. (AsFAD).



Figure 1. Phylogenetic analysis of *A. spinosa* L., *FAD* and *SAD* genes. Phylogenetic tree for *FAD* gene family was constructed by the "Neighbour-Joining" method using MEGA X with 500 bootstraps using orthologs from *Olea europaea* L., *Sesamum indium* L., *Theobroma cacao* L., *Linum usitatissimum* L., and *Arabidopsis thaliana* L.

Different clusters of tree are marked with different colors: Δ -12 desaturase (blue), Δ -15 desaturase (purple), stearoyl-ACP desaturase (green), and front-end desaturase (orange).







Figure 3. Identification of conserved motifs in *A. spinosa* L. *FADs*. Ten conserved motifs are presented in different colors. The identified motifs were detected by MEME online tool.

which contains only two His-boxes (1 and 2). All *AsFAD2* proteins contained three conserved His-boxes, HECGH, HRRHH, and HVAHH. However, *AsFAD3*/7 showed a variable second residue with the E residue replaced by D in His-Box1, whereas the third R and A residues were replaced by T and I residues in His-Box2 and His-Box3, respectively (Table 2). The members of the *AsSAD* family were quite divergent in residues in both His-Box1 and His-Box2 when compared with *AsFAD* members. They contained two conserved histidine boxes: EENRHG and DEKRHE (Table 2).

Cis elements in the promoters of AsFAD genes

We conducted additional research on cis-regulatory elements found in the promoter regions of *AsFADs*. Our findings revealed that the *AsFAD* gene family contains 79 cis-elements



Ten conserved motifs are presented in different colors. The identified motifs were detected by MEME online tool.

that can control gene expression in response to four categories: light, hormone-responsive, stress, and plant growth and development (Figure 7, Supplemental Material). The findings showed that the AsFADs promoters shared similar cis-regulatory components, such as CAAT-box, TATA-box, and MYC. Multiple cis-regulatory elements associated with plant hormone responses, including methyl jasmonate (MeJA)-, salicylic acid (SA)-, abscisic acid (ABA)-, gibberellin-, auxin-, and ethvlene-responsive elements, were identified in the AsFADs promoters, with ABA-responsive elements (28) within the highest number. In addition, all AsFADs promoters were found to contain light-responsive elements, with Box 4 being the most abundant element (48), present in all the promoters except for AsFAD2-d, AsFAD7-a, AsSAD1, AsSAD2, and AsSAD9. Several stress-responsive elements were present in the AsFADs promoters. Of these, the largest number (52) belonged to the droughtresponsive elements (MYC). In addition, certain cis-regulatory elements related to plant growth and development were discovered in AsFADs promoters, such as the A-box for specific day length, O2-site for zein metabolism, CCGTCC-box for meristem-specific gene activation, CCAAT-box for MYBHv1 binding, and AAGAA-motif for seed-specific expression.

Predicted 3D structures of AsFAD proteins

The 3D and TM structures of the *AsFAD*s were predicted by Phyre2 tool (Figure 8). The percentage of modeled residues ranged from 14% to 91% at >98% CI (Supplemental Table S2). The crystal structure of the human integral membrane stearoyl-CoA desaturase-2 was used for the protein model as the template structure for the *AsFAD2* subfamily, while the crystal structure of stearoyl-coenzyme A desaturase-1 was used for other *AsFAD* members. All members of the *AsSAD* subfamily have ribonucleotide reductase-like template structures. The modeled structure of *AsFAD*s showed two protein secondary structures: alpha helices and beta sheets. Common characteristics of all members of the *AsFAD* family include six TM helices (TM1-6; Figure 9).



The red rectangles represent conserved motifs (EENRHG, DEKRHE), which represent two His-boxes.

Designing and validation of primers

The details of the primer sets are summarized in Table 3, Supplemental Table S3. Nine *FAD* and nine *SAD* gene sequences, predicted from Argan genome, were analyzed to identify candidate primer sequences. Five primer pairs were designed for each gene, resulting in 90 primer pairs. Of these, 14 primers were selected based on the best characteristics, were commercially synthesized by Biologicio, and were successfully tested by PCR on three samples of genomic DNA extracted from Argan leaves. After numerous optimizations of the PCR amplification process, 11 primer pairs were found to amplify the expected genomic DNA in all samples (as shown in Figure 10). The expected product ranged from 197 (AsSAD7) to 1743 bp (AsSAD9).

Discussion

With the increasing number of plant genome sequences being generated, there are greater opportunities to explore the field of functional genomics and improve the accuracy of annotating newly discovered genes. Accordingly, genome-wide identification methods provide higher accuracy for characterizing multiple gene families, including the *FAD* gene family. To date, a large number of *FAD* homologs have been identified and characterized in several plant species, including 29 *FAD*s in soybean,¹⁵ 33 in sunflower,³⁵ 17 in Arabidopsis,³⁶ 36 in peanuts,³⁷



The red rectangles represent conserved motifs (H(X)2-4H, H(X)2-3HH, and H/Q(X)2-3HH), which represent three His-boxes

and 41 in Raymond cotton.³⁸ In the current study, we performed the first investigation of *FAD* in Argan and identified 18 putative members of the *FAD* gene family in *A spinosa* L. This number was consistent with the Kabuli chickpea genome, as predicted by Saini and Kumar,³⁶ and higher than that identified in Arabidopsis (17). The *SAD* subfamily in *A spinosa* L. has the highest number with nine members in comparison with other oil crop species such as *O europaea* L. and *S indium* L. Therefore, numerous species showed variety in their number of *FAD* gene families, as a result of species-specific process expansion induced by gene duplication events.³⁹ Moreover, this expansion may be associated with the large genome size of each species.⁴⁰

There are two distinct pathways in plants that lead to the biosynthesis of polyunsaturated fatty acids in the plastid and ER.⁴¹ According to our results, the subcellular localization of *AsFAD* and *AsSAD* is chloroplast/ER and chloroplast, respectively, which is in agreement with a previous study.^{9,42} Desaturation reactions are divided into two major classes: soluble *FAD* proteins and membrane-bound *FAD.*⁴ As per the findings of the phylogenetic analysis, four subfamilies of desaturases have been identified, namely Δ -12 desaturase (*FAD2/FAD6*), Δ -15 desaturase (*FAD3/FAD7*), *SAD*, and front-end desaturase (*FAD5*). This clustering of *AsFAD* proteins is in agreement with that observed in other plants, including *O europaea* L.,⁴³ *Triticum aestivum* L.,⁴⁴ and *A thaliana* L.⁴⁵ It could be suggested that the distribution of Argan *FAD* proteins throughout evolution is linked to their motif content, as the amino acid compositions within each cluster were similar.

TYPE	PROTEIN	HIS-BOX1		HIS-BOX2		HIS-BOX3	
		SEQUENCE	POSITION	SEQUENCE	POSITION	SEQUENCE	POSITION
ω-3/Δ-5	AsFAD3-a	HDCGH	151-155	HRTHH	187-191	HVIHH	354-358
	AsFAD7-a	HDCGH	168-172	HRTHH	204-208	HVIHH	371-375
	AsFAD3-b	HDCGH	128-132	HRTHH	164-168	HVIHH	332-336
	AsFAD7-b	HDCGH	175-179	HRTHH	211-215	HVIHH	378-382
ω-6/Δ-12	AsFAD2-a	HECGH	105-109	HRRHH	141-145	HVAHH	315-319
	AsFAD2-b	HECGH	107-112	HRRHH	143-147	HVAHH	317-321
	AsFAD2-c	HECGH	106-111	HRRHH	142-146	HVAHH	316-320
	AsFAD2-d	HDSGH	151-155	HNAHH	188-192	QVEHH	365-369
	AsFAD6	HDCAH	168-172	HDQHH	204-208	_	_
Δ-9	AsSAD1	EENRHG	111-116	DEKRHE	198-203	_	_
	AsSAD2	EENRHG	184-189	DEKRHE	270-275	_	_
	AsSAD3	EENRHG	184-189	DEKRHE	270-198	_	—
	AsSAD4	EENRHG	111-116	DEKRHE	197-202	_	_
	AsSAD5	EEKRHG	170-175	DEKRHE	256-261	_	_
	AsSAD6	EENRHG	206-211	DEKRHE	292-297	_	_
	AsSAD7	EEKRHG	170-175	DEKRHE	256-261	_	_
	AsSAD8	EENRHG	163-168	DEKRHE	249-254	_	_
	AsSAD9	EENRHG	177-182	DEKRHE	263-268	_	_

Table 2. The conserved histidine boxes of the FAD family proteins in Argania spinosa L.

SAD is the only identified soluble desaturase present in plants and is responsible for the conversion of stearic acid (18:0-ACP) to oleic acid (18:1 Δ 9-ACP).⁴⁶ Similarly, our results showed that A spinosa L. contains this type of gene, which could have the possibility of desaturation. The AsSAD subfamily was clustered close to the SAD protein from olive. On the other hand, the membrane-bound FAD known as FAD2/6 genes encoding Δ -12 desaturase/ ω -6 enzyme, which catalyzed the desaturation of oleic acid (18:1 Δ 9) to linoleic acid (18:2 Δ 9,12), The FAD3/7/8 genes encoding Δ -12 desaturase/ ω -6 enzyme, which performed the conversion of linoleic acid (18:2 Δ 9, 12) to linolenic acid (18:3 Δ 9, 12, 15), The FAD5 genes encoding frontend desaturase enzymes that convert palmitic acid (16:0) to palmitoleic acid (16:1) were grouped into the same clusters with an exact distinction in the phylogenetic tree.

The splicing phase and exon-intron structures play an important role in the evolution of gene families.⁴⁷ The two intron phases (0 and 1) in *AsFAD* were highly conserved, whereas intron phase 2 exhibited the lowest level of conservation. The same result for the splicing phase was observed in other plants, demonstrating that *FAD* genes have been highly conserved throughout evolution.¹⁵ Analysis of the structure of *AsFAD* genes revealed a similar intron-exon distribution pattern

within the AsFAD subfamily in terms of length. Eg, AsFAD2, AsFAD3, AsFAD6, and AsFAD7 are highly conserved. Motif analysis showed that all AsFAD genes contained motifs 1 and 2. The AsFAD3/7 subfamily had all motifs, except motif 10, which is why they were clustered into the same group. The AsFAD2 subfamily was clustered into one group because of the presence of similar motif content. The AsFAD6 and AsFAD2-d subfamilies had only two motifs. In contrast, the AsSAD subfamily contained all six motifs; thus, it was grouped in a separate cluster from the AsFAD subfamily. AsFAD and AsSAD showed the presence of FA_desaturase and FA_desaturase_2 domains, respectively. Three histidine boxes were observed in the AsFAD subfamily, with the exception of AsFAD6, which had two Hisboxes. AsFAD proteins contained a consensus sequence of H/Q(X)2HH for their third His-box, which was situated at their carboxy terminus. This sequence was found in FAD protein sequences across different plant species, suggesting a relationship between them.¹⁶ With the exception of AsFAD2-d, the number of residues between the first and second His-boxes remained consistently conserved among members of the AsFAD subfamily. Histidine boxes are significantly involved in the development of the di-iron center, which is responsible for both oxygen activation and substrate oxidation.²³ The histidine-rich



Figure 7. Distribution of cis-acting regulatory elements in the *AsFAD* promoter.

regions of $F\!AD$ proteins exhibit a high degree of conservation in their amino acid sequence.⁴⁸

To enhance our comprehension of the transcriptional regulation and potential functional roles of *AsFAD* genes, we conducted a gene promoter analysis, which showed that most of the *AsFAD* genes contain a cis-element that plays a role in responding to different types of stress within their promoter regions, including light, hormone-responsive (such as ethylene, ABA, and SA), stress (such as defense, wound, and cold/dehydration), and plant growth and development. This suggests that *AsFAD* genes have various functions in different stress regulatory networks.⁴⁹ Regulation of signaling pathways in response to abiotic stress is crucial, and transcription factors are key players in this process.⁵⁰ This is achieved by binding to the promoters of specific genes and controlling their activation or inhibition. Desaturases are known to respond to various types of stress, such as low temperatures, where the expression of *FAD* genes is modified, leading to alterations in the fluidity of membrane



Figure 8. Predicted 3D structures of *A. spinosa* L. *FAD* proteins. The 3D structure model and the template of *AsFAD*s were performed by the Phyre2 Software server (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index).

lipids and ultimately enhancing the ability of plants to withstand cold conditions. $^{51}\,$

Desaturation reactions by *FADs* are necessary for the biosynthesis of fatty acids because they lead to the formation of double bonds in the hydrocarbon chains of fatty acids. The presence of unsaturated fatty acids in vegetable oils is directly linked to their nutritional value and resistance to oxidation. Discovering *AsFAD* genes will facilitate the creation of enhanced *A spinosa* L. plants, which will not only have improved nutritional value but also greater resistance to diverse environmental factors. The current study contributes to the understanding of the potential functions and evolution of the *AsFAD*



Figure 9. The predicted transmembrane topology of A. spinosa L. FAD from Phyre2.



Figure 10. An example of PCR amplification of DNA of three samples (1, 2, 3) extracted from argan leaves by three specific primers (*FAD2-a*, *FAD2-b*, and *FAD2-c*) on 2% agarose gel. M: 100bp DNA marker. PCR indicates polymerase chain reaction.

genes. Dar et al⁵² clarified the importance and key role of *FAD2* in cold and salt tolerance, plant development, and fatty acid biosynthesis in plants. Research has demonstrated that the linolenic acid content of plants is closely linked to the activity of the *FAD3* gene, and plays a vital role in regulating plant fatty acid composition, which directly affects oil quality. Furthermore, Zhang et al⁵³ showed that *MsFAD3* is involved in the synthesis of α -linolenic acid and that the α -linolenic acid content was significantly increased in *MsFAD3* overexpression transgenic alfalfa lines. In our study, *AsSAD* presented the highest number of gene copies (9), which may be related to the highest percentage of oleic acid in argan oil.⁵⁴

Recognizing the central role of *FAD* genes in plant development, stress response, quality oil, and other physiological processes, it is important to develop molecular markers (candidate primers) for *FAD* and *SAD* genes in *A spinosa* L. These markers can be useful for identifying and selecting desirable traits such as increased oil content, improved nutritional quality, and enhanced stress tolerance, all of which are critical for sustainable agriculture and food security. To achieve this, 14 specific primer pairs were developed for both *FAD* and *SAD* genes in the Argan genome, which were validated to identify the

presence of these genes. Most of these primer pairs (11) showed reliable amplification of the expected size. Based on these results, the identified candidate primers could be considered as potential markers for the selection of valuable ecotypes in Argan. Furthermore, it would be helpful to conduct studies on the functional genomics of *A spinosa* L., such as gene expression studies, to improve the quality of its oil.

Conclusion

In conclusion, we presented a comprehensive in silico study of *FAD*-encoding genes in the Argan genome. To the best of our knowledge, this is the first genome-wide study on the Argan *FAD* gene family. Overall, 18 putative *FAD* genes (*AsFAD* and *AsSAD*) in *A spinosa* L. were phylogenetically clustered into two main classes (membrane-bound desaturases and soluble desaturases). In addition, each subfamily demonstrated a high degree of consistency in subcellular localization, motif composition, and exon-intron organization. Eleven pairs of primers were designed, tested, and validated on genomic DNA extracted from three accessions of argan leaves. Our comprehensive analysis of *FAD* genes in this study not only screened candidate genes for functional validation but also provided resources and references to enhance the abiotic stress and oil quality of *A. spinosa* L. oil.

Acknowledgements

The authors acknowledge Mohammed V University, Faculty of Science, Rabat for their support.

Author Contributions

A.E. conceived, designed, and performed the experiments and wrote original draft. R.K. reviewed the draft and finalized the manuscript. A.M. contributed to the analysis of samples in the laboratory. F.M.F. designed the study and examined the document. B.B. developed conceptual ideas and edited the revised manuscript. All authors read and approved the final manuscript.

PRIMER NAME	PRIMER SEQUENCE	PCR PRODUCT	ANNEALING TM
AspSAD1 p1	5' AACTAAGGGCAAGGGCAAAT 3' 3' TCATGTCTACGCGTCCAGAG 5'	243	55.6
AspSAD2 p5	5' TTCTTCGAGAAAGCCAAGGA 3' 3' CATGTCGGTTCTCTTCAGCA 5'	214	55.25
AspSAD4 p2	5' GGAAAGGGCAAAGGAGATTC 3' 3' GGTCACCATGCCTGTTTTCT 5'	194	53.3
AspSAD6 p2	5' GCTGCGAGAAAGATCAAAGG 3' 3' GCTTCATGTCAACTCGTCCA 5'	245	55.4
AspSAD7 p2	5' ATTCCATGCCACCTGAAAAG 3' 3' CAAACAATCGTCCGGAATCT 5'	197	53.1
AspSAD9 p1	5' AGCCTCAATCCTTGGGCTAT 3' 3' AAGCCTCGCTGTGTTTCTGT 5'	246	56.3
AspFAD2-a p2	3' CAAGTCCCAAATCTCGTGGT 5' 5' CCGACATCCGAAACGTAGAT 3'	210	54.4
AspFAD2-b p1	3' GCAGTAAAAGGGCTTGCTTG 5' 5' GAGGGATCGTGGAGAACAAA 3'	247	54.5
AspFAD2-c p2	3' GACTTCGCCATCGTTTTCAT 5' 5' GTAAGGGACGAGGAGGGAAG 3'	228	54.7
AspFAD3-a p2	5' GAGCTGCTATTCCCAAGCAC 3' 3' TGCAAGAGATGTCCAACAGC 5'	250	55.95
AspFAD3-b p5	5' AGGGGAGGGCTTACAACAGT 3' 3' AAAGTGGTCTTGCCTGATGC 5'	237	55.3

Table 3. The 11 pair primers successfully amplified the expected size of gDNA in A. spinosa L.

Abbreviation: PCR, polymerase chain reaction; TM, melting temperature

SUPPLEMENTAL MATERIAL

Supplemental material for this article is available online.

REFERENCES

- Guillaume D, Charrouf Z. Argan oil and other argan products: use in dermocosmetology. Eur J Lipid Sci Technol. 2011;113:403-408.
- El Kharrassi Y, Maata N, Mazri MA, et al. Chemical and phytochemical characterizations of argan oil (*Argania spinosa* L. skeels), olive oil (*Olea europaea* L. cv. *Moroccan picholine*), cactus pear (*Opuntia megacantha salm-dyck*) seed oil and cactus cladode essential oil. J Food Meas Charact. 2018;12:747-754.
- Vingrys AJ, Armitage JA, Weisinger HS, Bui BV, Sinclair AJ, Weisinger RS. The role of omega-3 polyunsaturated fatty acids in retinal function. In: Mostofsky DI, Yehuda S, Salem N, eds. *Fatty Acids: Physiological and Behavioural Functions*. Totowa, NJ: Humana Press; 2001:193-217.
- Cheng C, Liu F, Sun X, et al. Genome-wide identification of FAD gene family and their contributions to the temperature stresses and mutualistic and parasitic fungi colonization responses in banana. *Int J Biol Macromol.* 2022;204:661-676.
- Simões T, Ferreira J, Lemos MF, et al. Argan oil as a rich source of linoleic fatty acid for dietetic structured lipids production. *Life*. 2021;11:1114.
- Gharby S, Charrouf Z. Argan oil: chemical composition, extraction process, and quality control. *Front Nutr.* 2022;8:1251.
- Sharma A, Chauhan RS. In silico identification and comparative genomics of candidate genes involved in biosynthesis and accumulation of seed oil in plants. *Comp Funct Genomics*. 2012;2012:914843.
- Nakamura MT, Nara TY. Structure, function, and dietary regulation of Δ6, Δ5, and Δ9 desaturases. Annu Rev Nutr. 2004;24:345-376.
- Celik Altunoglu Y, Unel NM, Baloglu MC, Ulu F, Can TH, Cetinkaya R. Comparative identification and evolutionary relationship of fatty acid desaturase (FAD) genes in some oil crops: the sunflower model for evaluation of gene expression pattern under drought stress. *Biotechnol Biotechnol Equip*. 2018;32:846-857.
- Chen N, Yang Q, Pan L, et al. Identification of 30 MYB transcription factor genes and analysis of their expression during abiotic stress in peanut (*Arachis hypogaea* L.). Gene. 2014;533:332-345.

- Sperling P, Ternes P, Zank TK, Heinz E. The evolution of desaturases. Prostaglandins Leukot Essent Fatty Acids. 2003;68:73-95.
- glandins Leukot Essent Fatty Acids. 2003;68:73-95.
 12. Chen G, Xing ZK, Pan WL, et al. Cloning of a novel stearoyl-acyl desaturase gene from white ash (*Fraxinus americana*) and evolution analysis with those from other plants. *Afr J Biotechnol.* 2011;10:18185-18193.
- Wallis JG, Browse J. Mutants of *Arabidopsis* reveal many roles for membrane lipids. *Prog Lipid Res*, 2002;41:254-278.
- Xue Y, Chen B, Wang R, Win AN, Li J, Chai Y. Genome-wide survey and characterization of fatty acid desaturase gene family in *Brassica napus* and its parental species. *Appl Biochem Biotechnol.* 2018;184:582-598.
- Chi X, Yang Q, Lu Y, et al. Genome-wide analysis of fatty acid desaturases in soybean (*Glycine max*). *Plant Mol Biol Report*. 2011;29:769-783.
- Díaz ML, Cuppari S, Soresi D, Carrera A. In silico analysis of fatty acid desaturase genes and proteins in grasses. *Appl Biochem Biotechnol.* 2018;184:484-499.
- Khodakovskaya M, McAvoy R, Peters J, Wu H, Li Y. Enhanced cold tolerance in transgenic tobacco expressing a chloroplast ømega-3 fatty acid desaturase gene under the control of a cold-inducible promoter. *Planta*. 2006;223:1090-1100.
- Zhang J, Liu H, Sun J, et al. Arabidopsis fatty acid desaturase FAD2 is required for salt tolerance during seed germination and early seedling growth. *PLoS ONE*. 2012;7:e30355.
- Zhang JT, Zhu JQ, Zhu Q, Liu H, Gao XS, Zhang HX. Fatty acid desaturase-6 (Fad6) is required for salt tolerance in *Arabidopsis thaliana*. *Biochem Biophys Res Commun*. 2009;390:469-474. doi:10.1016/j.bbrc.2009.09.095.
- Wickramanayake JS, Goss JA, Zou M, Goggin FL. Loss of function of fatty acid desaturase 7 in tomato enhances photosynthetic carbon fixation efficiency. *Front Plant Sci.* 2020;11:932.
- Zhang M, Barg R, Yin M, et al. Modulated fatty acid desaturation via overexpression of two distinct ømega-3 desaturases differentially alters tolerance to various abiotic stresses in transgenic tobacco cells and plants. *Plant J.* 2005;44: 361-371.
- Niu E, Gao S, Hu W, et al. Genome-wide identification and functional differentiation of fatty acid desaturase genes in *Olea europaea L. Plants.* 2022;11:1415.
- Dong CJ, Cao N, Zhang ZG, Shang QM. Characterization of the fatty acid desaturase genes in cucumber: structure, phylogeny, and expression patterns. *PLoS ONE*. 2016;11:e0149917.

- Artimo P, Jonnalagedda M, Arnold K, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012;40:W597-W603.
- Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins Struct Funct Bioinforma. 2006;64:643-651.
- Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. DeepLoc: prediction of protein subcellular localization using deep learning. *Bio-informatics*. 2017;33:3387-3395.
- Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31:1296-1297.
- Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 2006;34:W369-W373.
- Finn RD, Bateman A, Clements J, et al. Pfam: the protein families database. Nucleic Acids Res. 2014;42:D222-D230.
- Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870-1874.
- Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783-791.
- Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30:325-327.
- Basyuni M, Wati R, Sulistiyono N, et al. Protein modelling of triterpene synthase genes from mangrove plants using Phyre2 and Swiss-model. J Phys Conf Ser. 2018;978:012095.
- Martínez-Rivas JM, Sperling P, Lühs W, Heinz E. Spatial and temporal regulation of three different microsomal oleate desaturase genes (FAD2) from normal-type and high-oleic varieties of sunflower (Helianthus annuus L.). *Mol Breed*. 2001;8:159-168.
- Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR. Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Ara-bidopsis. Science*. 1992;258:1353-1355.
- Peng Z, Ruan J, Tian H, et al. The family of peanut fatty acid desaturase genes and a functional analysis of four ømega-3 AhFAD3 members. *Plant Mol Biol Report*. 2020;38:209-221.
- Feng J, Dong Y, Liu W, et al. Genome-wide identification of membrane-bound fatty acid desaturase genes in *Gossypium hirsutum* and their expressions during abiotic stress. *Sci Rep.* 2017;7:45711.
- Chalhoub B, Denoeud F, Liu S, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science*. 2014;345:950-953.
- Ma J, Yang Y, Luo W, et al. Genome-wide identification and analysis of the MADS-box gene family in bread wheat (*Triticum aestivum L.*). *PLoS ONE*. 2017;12:e0181443.

- Sato N, Moriyama T. Genomic and biochemical analysis of lipid biosynthesis in the unicellular rhodophyte *Cyanidioschyzon merolae*: lack of a plastidic desaturation pathway results in the coupled pathway of galactolipid synthesis. *Eukaryot Cell*. 2007;6:1006-1017.
- Shang X, Cheng C, Ding J, Guo W. Identification of candidate genes from the SAD gene family in cotton for determination of cottonseed oil composition. *Mol Genet Genomics*. 2017;292:173-186.
- Contreras C, Mariotti R, Mousavi S, et al. Characterization and validation of olive FAD and SAD gene families: expression analysis in different tissues and during fruit development. *Mol Biol Rep.* 2020;47:4345-4355.
- Hajiahmadi Z, Abedi A, Wei H, et al. Identification, evolution, expression, and docking studies of fatty acid desaturase genes in wheat (*Triticum aestivum* L.). *BMC Genomics*. 2020;21:1-20.
- Xu L, Zeng W, Li J, et al. Characteristics of membrane-bound fatty acid desaturase (FAD) genes in *Brassica napus* L. and their expressions under different cadmium and salinity stresses. *Environ Exp Bot.* 2019;162:144-156.
- Sun R, Gao L, Yu X, Zheng Y, Li D, Wang X. Identification of a Δ12 fatty acid desaturase from oil palm (*Elaeis guineensis* Jacq.) involved in the biosynthesis of linoleic acid by heterologous expression in *Saccharomyces cerevisiae*. Gene. 2016;591:21-26.
- Eshkiki EM, Hajiahmadi Z, Abedi A, Kordrostami M, Jacquard C. In silico analyses of autophagy-related genes in rapeseed (*Brassica napus* L.) under different abiotic stresses and in various tissues. *Plants*. 2020;9:1393.
- Alonso DL, Garcia-Maroto F, Rodriguez-Ruiz J, Garrido JA, Vilches MA. Evolution of the membrane-bound fatty acid desaturases. *Biochem Syst Ecol.* 2003;31:1111-1124.
- Gai W, Sun H, Hu Y, et al. Genome-wide identification of membrane-bound fatty acid desaturase genes in three peanut species and their expression in *Arachis hypogaea* during drought stress. *Genes.* 2022;13:1718.
- Liu J, Wang F, Yu G, et al. Functional analysis of the maize C-repeat/DRE motif-binding transcription factor CBF3 promoter in response to abiotic stress. *Int J Mol Sci.* 2015;16:12131-12146.
- He M, Ding NZ. Plant unsaturated fatty acids: multiple roles in stress response. Front Plant Sci. 2020;11:562785.
- Dar AA, Choudhury AR, Kancharla PK, Arumugam N. The FAD2 gene in plants: occurrence, regulation, and role. *Front Plant Sci.* 2017;8:1789.
- Zhang Z, Jin X, Liu Z, Zhang J, Liu W. Genome-wide identification of *FAD* gene family and functional analysis of *MsFAD3.1* involved in the accumulation of α-linolenic acid in alfalfa. *Crop Sci.* 2021;61:566-579.
- Azizi SE, Dalli M, Mzabri I, Berrichi A, Gseyra N. Chemical characterization of oils produced by some native and introduced genotypes of argan tree in eastern Morocco using HPLC-DAD/GC-MS, and the evaluation of their physicochemical parameters. OCL. 2022;29:14.