

# Friedelin in *Maytenus ilicifolia* Is Produced by Friedelin Synthase Isoforms

Thaís B. Alves <sup>1</sup>, Tatiana M. Souza-Moreira <sup>2</sup>, Sandro R. Valentini <sup>2</sup>, Cleslei F. Zanelli <sup>2,\*</sup> and Maysa Furlan <sup>1</sup>

<sup>1</sup> Instituto de Química, Univ. Estadual Paulista-UNESP, Rua Prof. Francisco Degni, 55, Quitandinha, Araraquara, SP 14800-060, Brazil; thais.barboni@gmail.com (T.B.A.); mayfurlan@gmail.com (M.F.)

<sup>2</sup> Faculdade de Ciências Farmacêuticas, Univ. Estadual Paulista-UNESP, Rod. Araraquara-Jaú km 1, Araraquara, SP 14801-903, Brazil; souzatom@gmail.com (T.M.S.-M.); valentsr@fcar.unesp.br (S.R.V.)

\* Correspondence: zanellicf@fcar.unesp.br; Tel.: +55-16-3301-6996

**Table S1.** Comparison of variant sequences among *MiFRS* sequences.

Sequence	Region	Allele	Synonymous	Codon position	Protein residue	Amino acid position
<i>MiFRS</i>	Exon 1	24C	yes	3	Asp	8
		81G	yes	3	Arg	27
		102C	yes	3	Pro	34
		115C	non	1	Pro	39
		143C	non	2	Ala	48
		195C	yes	3	Leu	65
	Exon 5	1295G	non	1	Glu	309
	Exon 15	3703T	non	2	Ile	768
<i>MiFRS2</i>	Exon 1	24T	yes	3	Asp	8
		81G	yes	3	Arg	27
		102T	yes	3	Pro	34
		115T	non	1	Ser	39
		143C	non	2	Ala	48
		195A	yes	3	Leu	65
	Exon 5	1295A	non	1	Lys	309
	Exon 15	3703A	non	2	Lys	768
<i>MiFRS3</i>	Exon 1	24T	yes	3	Asp	8
		81A	yes	3	Arg	27
		102T	yes	3	Pro	34

		115T	non	1	Ser	39
		143T	non	2	Val	48
		195A	yes	3	Leu	65
	Exon 5	1295A	non	1	Lys	309
	Exon 15	3703T	non	2	Ile	768
<b>MiFRS4</b>	Exon 1	24T	yes	3	Asp	8
		81A	yes	3	Arg	27
		102T	yes	3	Pro	34
		115T	non	1	Ser	39
		143T	non	2	Val	48
		195A	yes	3	Leu	65
	Exon 5	1295A	non	1	Lys	309
	Exon 15	3703T	non	2	Ile	768

**Table S2.** The sequences used in the phylogenetic analysis with their GenBank accession number, species and function.

Identification	Accession number	Species	Function
LUP1	GenBank: AEE36187.1	<i>Arabidopsis thaliana</i>	lupeol synthase 1
LUP2	GenBank: AEE36185.1	<i>Arabidopsis thaliana</i>	lupeol synthase 2
CAS1	GenBank: AAC04931.1	<i>Arabidopsis thaliana</i>	cycloartenol synthase
BPY	GenBank: BAB83088.1	<i>Betula platyphylla</i>	$\beta$ -amyrin synthase
BPW	GenBank: BAB83087.1	<i>Betula platyphylla</i>	lupeol synthase
BgLUS	UniProtKB: 8CDT3.1	<i>Bruguiera gymnorhiza</i>	lupeol synthase
EtAS	GenBank: BAE43642.1	<i>Euphorbia tirucalli</i>	$\beta$ -amyrin synthase
GgbAS1	GenBank: BAA89815.1	<i>Glycyrrhiza glabra</i>	$\beta$ -amyrin synthase
KdFRS	GenBank: ADK35125.1	<i>Kalanchoe daigremontiana</i>	friedelin synthase
KdLUS	GenBank: ADK35126.1	<i>Kalanchoe daigremontiana</i>	lupeol synthase
KdGLS	GenBank: ADK35124.1	<i>Kalanchoe daigremontiana</i>	glutinol synthase
KdTAS	GenBank: ADK35123.1	<i>Kalanchoe daigremontiana</i>	taraxerol synthase

KcMS	GenBank: BAF35580.1	<i>Kandelia candel</i>	multifunctional triterpene synthase
AMY2	GenBank: AAO33580.1	<i>Lotus japonicus</i>	multifunctional $\beta$ -amyrin synthase
OSC3	GenBank: BAE53430.1	<i>Lotus japonicus</i>	lupeol synthase
LcIMS1	UniProtKB: Q948R6.1	<i>Luffa cylindrica</i>	isomultiflorenol synthase
GgLUS1	UniProtKB: Q764T8.1	<i>Glycyrrhiza glabra</i>	lupeol synthase
OSC1	NCBI Reference Sequence: NP_001280946.1	<i>Malus domestica</i>	$\beta$ -amyrin synthase
MiFRS	GenBank: APG38073.1	<i>Maytenus ilicifolia</i>	friedelin synthase
MiFRS2	Under submission	<i>Maytenus ilicifolia</i>	friedelin synthase 2
MiFRS3	Under submission	<i>Maytenus ilicifolia</i>	friedelin synthase 3
MiFRS4	Under submission	<i>Maytenus ilicifolia</i>	friedelin synthase 4
AMY1	GenBank: AAO33578.1	<i>Medicago truncatula</i>	$\beta$ -amyrin synthase
OEW	GenBank: BAA86930.1	<i>Olea europaea</i>	lupeol synthase
PNY1	GenBank: BAA33461.1	<i>Panax ginseng</i>	$\beta$ -amyrin synthase
PNY2	GenBank: BAA33722.1	<i>Panax ginseng</i>	$\beta$ -amyrin synthase
PSM	GenBank: BAA97559.1	<i>Pisum sativum</i>	mixed-amyrin synthase
PSY	GenBank: BAA97558.1	<i>Pisum sativum</i>	$\beta$ -amyrin synthase
PdFRS	GenBank: ART66198.1	<i>Populus davidiana</i>	friedelin synthase
RsM1	GenBank: BAF80441.1	<i>Rhizophora stylosa</i>	multifunctional triterpene synthase
RcLUS	NCBI Reference Sequence: NP_001310684.1	<i>Ricinus communis</i>	lupeol synthase
TRW	GenBank: BAA86932.1	<i>Taraxacum officinale</i>	lupeol synthase
BPX1	GenBank: BAB83085.1	<i>Betula platyphylla</i>	cycloartenol synthase
GgCAS1	GenBank: BAA76902.1	<i>Glycyrrhiza glabra</i>	cycloartenol synthase

Table S2. Cont.

KdCAS	GenBank: ADK35127.1	<i>Kalanchoe daigremontiana</i>	cycloartenol synthase
OSC5	GenBank: BAE53431.1	<i>Lotus japonicus</i>	cycloartenol synthase
LcCAS1	UniProtKB: Q9SLP9.1	<i>Luffa cylindrica</i>	cycloartenol synthase
MiCAS1	GenBank: APG38074.1	<i>Maytenus ilicifolia</i>	cycloartenol synthase 1
PNX	GenBank: BAA33460.1	<i>Panax ginseng</i>	cycloartenol synthase
PSX	GenBank: BAA23533.1	<i>Pisum sativum</i>	cycloartenol synthase

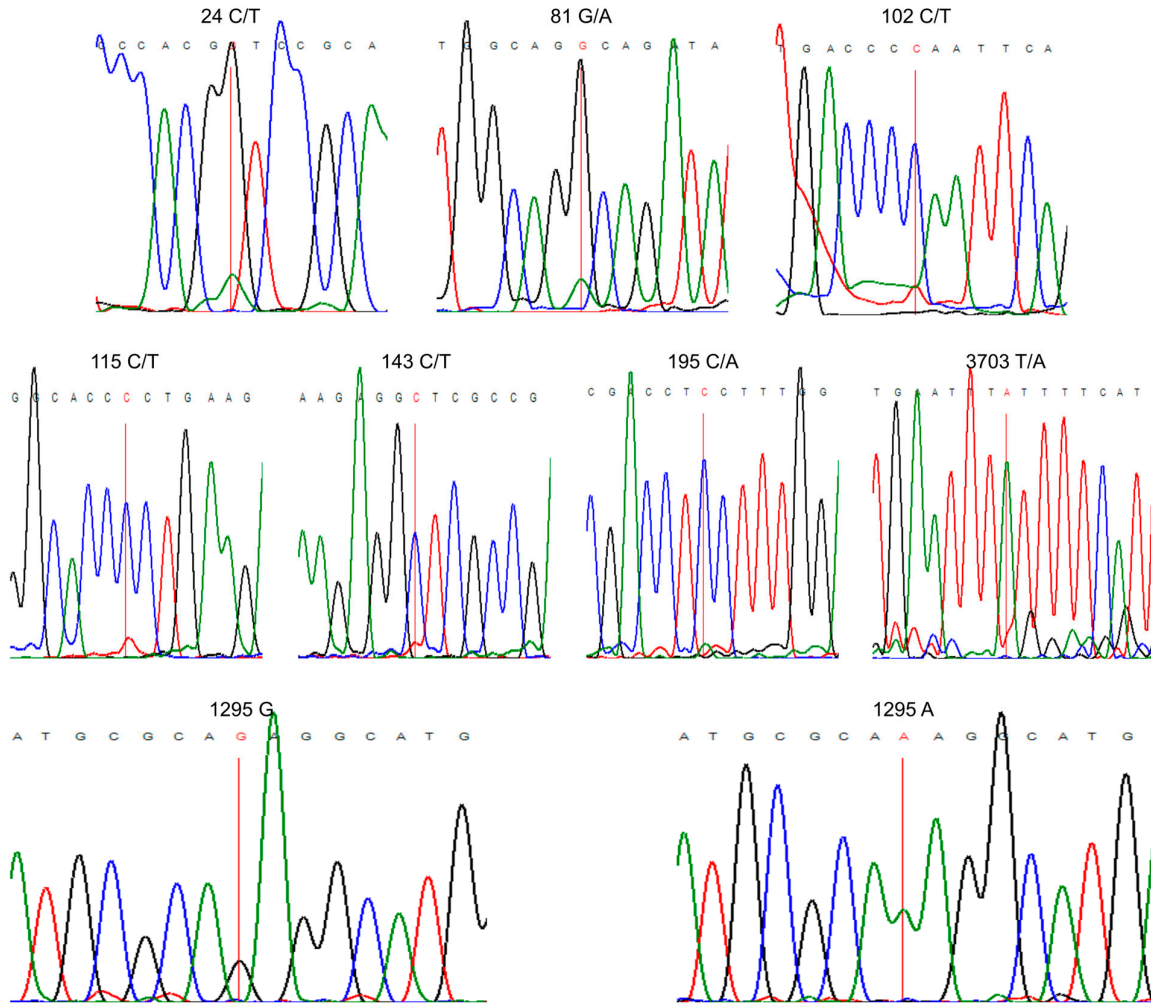
Table S3. Predicted exon-intron flanking consensus interfaces in MiFRS gene sequence.

Donor splice sites, direct strand, 5'exon^3'intron	Acceptor splice sites, direct strand, 5'intron^3'exon
GATGATGCAG^GTA	
TCCTCCATT^GTGAGTATTT	AAA
TTGGTTTTCG^GTGCGATAAA	
ATACATCCAG^GTA	ATCA
ATGCGCAGAG^GCATGGATTC	TACCTTGCAG^GAGGATCTGT
CGTTGAAAAG^GTATCAAATA	TATCGAGCAG^GCGCTGTGTA
GAAAGTGGTG^GTAAGTAGAG	TCATGTGCAG^AGTTTTGGCA
AGCCTTAAAG^GTAGTGAATT	TATTTTACAG^TGCTGCCTGC
CCTCCAAGAG^GTAAAATTGA	GTCATTGCAG^CCTAAA
ATGGATAGAG^GTA	CCTTTTATAG^ATGCTCAACC
ATGGCTCGTG^GTACATTCTC	GTTTATCTAG^GTATGGGAAC
CCCAAGAAAG^GTGCATTTTT	GGTGCTACAG^AAATATACGC
TGCTGGCCAG^GTGAAA	CCCAAATAG^GCTGAGAGAG
CCCTCAGCAG^GTA	

Table S4. Primers used in this study.

Primer	Sequence 5' – 3'	Function
VZO1381	CG <u>CGGATCC</u> ATCATGTGGAAGATAAAGATTGC	Forward for full-length ORF amplification. <i>Bam</i> HI cleavage site. Also used for gene amplification and sequencing.
VZO1382	CG <u>GAAATTC</u> CGTGTAAGTGGACATATCATATGA	Reverse for full-length ORF amplification. <i>Eco</i> RI cleavage site. Also used for gene amplification and sequencing.
M13F	GTAAAACGACGGCCAGT	ORF sequencing
M13R	CAGTATCGACAAAGGAC	ORF sequencing
VZO1473	GGGACACAGCACAGTGTTT	ORF and gene sequencing
VZO1471	CTCACAGATGAAGTAGCTCCTACTC	ORF sequencing
VZO1472	CCGAGTGCAAACATTGTAC	ORF sequencing
VZO1134	GGAGGGCTAGCAGCAGCTGG	ORF and gene sequencing
VZO1950	GGCATGTAAACCATCC	Gene sequencing
VZO1951	GGTTGCTGAGGATGG	Gene sequencing
VZO1952	GGAGTTTCTTGAGAAATATTGTG	Gene sequencing
VZO2067	CTGCTCCGACCTCCTTTGG	Gene sequencing

Endonuclease restriction sites are underlined.



**Figure S1.** Electropherogram of SNPs detected in the gene of *MiFRS*. Electropherograms corresponding to nucleotide positions 24, 81, 102, 115, 143, 195, 3703 represent SNPs in exon 1 and 15. The position is indicated above each electropherogram, which were generated by sequencing of the gene PCR product. SNP in position 1295 correspond to SNP in exon 5, whose PCR product was cloned into plasmids and then sequenced individually. Therefore, it is possible to see one electropherogram for each polymorphism. Images were generated by the QSVAnalyser tool. Electropherogram for SNP 24 C/T and 3703 T/A were observed by using reverse primer for sequencing.

BPW	27	FDFDAG-TPEERAEEVRE	306	ARSTVAKEDLY	752	VLFA-----
OSC3	27	FDENAG-TPEEHEEEMRQ	306	ARNTVAKEDLY	752	VLCA-----
GgLUS1	27	FDENAG-TPEEHAEEELRR	306	ARNTVAKEDLY	752	VLCAHSY----
OEW	28	FDFEAG-TPEERVEVERLRE	309	ARNTCAKEDLY	753	TLHAQT-----
TRW	29	FDFDAG-TEEERAEEKIRL	307	ARNTCAKEDLY	755	VQNI-----
FRS	29	YDENAG-TPEEREQVEARR	305	VRHCAEEDLY	752	VPLPSKKL----
MiFRS	32	FDENSG-TPEELAEEARR	308	MRHCAEEDLY	755	VPLPSKGNMAMKNSA--
MiFRS2	32	FDENSG-TSEELAEEARR	308	MRHCAKEDLY	755	VPLPSKGNMAMKNSA--
MiFRS3	32	FDENSG-TSEELAEEVRR	308	MRHCAKEDLY	755	VPLPSKGNMAMKNSA--
MiFRS4	32	FDENSG-TSEELAEEVRR	308	MRHCAKEDLY	755	VPLPSKGNMAMKNSA--
OSC1	29	FDFDAG-TEEERAEEVARE	308	VRHCAEDNY	752	VSLPIKKIA-----
RsM2	31	FDFDGG-TPEERDOVEARQ	310	ARHCAKEDLY	757	VLGLTSKAHSSAVME---
LcIMS1	29	FDENAG-TPEERAEEELRH	305	VRHCAEDLY	752	VPLPSKKK-----
LUP1	30	FDHKAG-SPEERAEEARR	306	SRRYAKEDLY	753	VFVN-----
LUP2	30	FDEKAG-TPEERAEEARR	309	ARRCAKEDLY	756	AFATHQDL-----
RcLUS	29	FDENAG-TPEEQAEVEARQ	308	VRHCAKEDNY	754	VLFPSAGFGFGFTNNL---
BgLUS	29	FDFEAG-TPEERAQVEARE	308	MRHCAKEDNY	754	FPLPSKND-----
KcMS	29	FDFEAG-TPEERAQVEARQ	308	MRHCAKEDNH	754	FPLPSKND-----
KdFRS	29	FDFQAT-TEQQLAKVEARL	308	VRHVCKEDTY	755	CNPNSEAISKPSK-----
KdLUS	29	FDFQAT-TEQQLAKVEARL	308	VRHVCKEDLY	755	LQLHSEATKMV-----
KdGLS	29	FDFQAT-TEQQLAKVEARL	308	VRHVCKEDLY	755	CNLSSEAISKPSK-----
PNY1	30	FDEDYVAEPESELEVEQVRR	310	TRRVCAKEDLY	756	VPLPSLGT-----
PNY2	29	FDFDYG-TPEERAEEVARR	308	VRHCAKEDLY	754	VRLPSKSV-----
EtAS	29	FDFQPP-TPEELAQVQCARL	308	TRHCAHEDLY	755	VPLPSTTL-----
RsM1	29	FDFDAG-TPEERAKAEARQ	308	NCHCAPEDLY	755	VPLPS-----
BPY	29	FDFQAG-SPEERAEEVARR	308	VRHCAKEDLY	755	VPLPLGKNLNQVNCIGQS
KdTAS	48	FDFEAG-TPEERAEEVARR	327	VRHCAKEDLY	774	LPLPLR-----
PSM	29	FDFDAG-TSCEKAQVEARQ	308	TRHCAKEDLY	755	ILLASPAVAI-----
AMY2	29	YDEDAG-TPEERAQVEARQ	308	ARHCAKEDLY	755	VPLPSIAV-----
GgbAS1	29	YDFDGG-TPEERAQVEARL	308	ARHCAKEDLY	755	VPLPSTPVCLT-----
PSY	29	YDFEAG-SPEERAQVEARR	308	TRHCAKEDLY	755	VPLP-----
AMY1	29	YDFEAG-SPEERAQVEARR	308	SRHCAKEDLY	755	VPLPSTAV-----

**Figure S2.** Multiple global alignment with diverse pentacyclic triterpene synthases. It is interesting to note the conservation of a Pro39 residue, an aliphatic residue at position 48, as Ala or Val and Lys309.

BPW	478	WQVSDCTAEGLK
OSC3	478	WQVSDCTAEGLK
GgLUS1	478	WQVSDCTAEGLK
OEW	479	WQVSDCTAEGLK
TRW	481	WQVSDCTAEGLK
FRS	478	WQLSDCTAEALK
MiFRS	480	WQLSDCTAEALK
MiFRS2	480	WQLSDCTAEALK
MiFRS3	480	WQLSDCTAEALK
MiFRS4	480	WQLSDCTAEALK
OSC1	480	WQVSDCTAEALR
RsM2	483	VIALDCTAESLM
LcIMS1	478	WQVSDCTAENLK
LUP1	479	WQVSDCTAEALK
LUP2	482	WQVSDCTAEALK
RcLUS	480	WQVSDCTAESLK
BgLUS	480	WQVSDCTAESLK
KcMS	480	WQVSDCTAESLK
KdFRS	481	WQLSDCTAEGLK
KdLUS	481	WQVSDCTAEGLK
KdGLS	481	WQLSDCTAEGLK
PNY1	482	WQVSDCTAEGLK
PNY2	480	WQVSDCTAEALK
EtAS	481	WQVSDCTAEGLK
RsM1	481	WQVSDCTAEGLK
BPY	481	WQVSDCTAEGLK
KdTAS	500	WQVSDCTAEGLK
PSM	481	WQVSDCTAESLK
AMY2	481	WQVSDCTAEGFK
GgbAS1	481	WQVSDCTAEGLK
PSY	481	WQVSDCTAEGLK
AMY1	481	WQVSDCTAEGLK

**Figure S3.** Leucine residue close to catalytic aspartate is conserved only in friedelin synthases. Multiple global alignment with diverse pentacyclic triterpene synthases shows the presence of a valine at the corresponding position in most of the enzymes, but not in the ones producing friedelin.



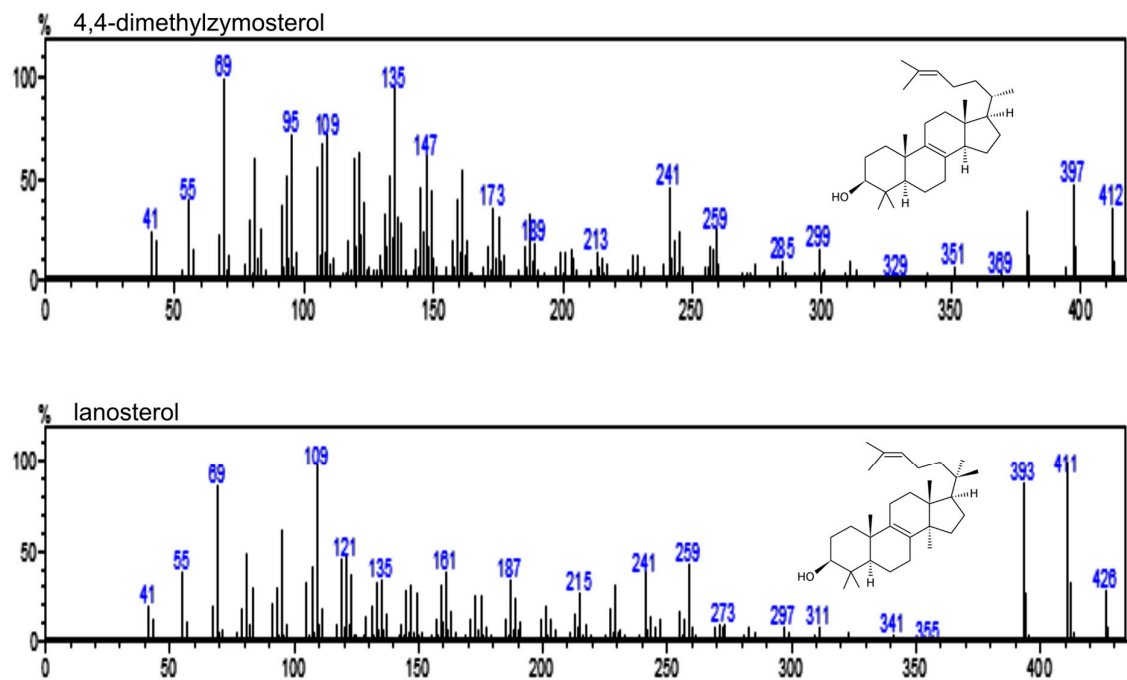


Figure S4. Mass spectra of 4,4-dimethylzymosterol and lanosterol.