Plant Gene Register

Genomic Nucleotide Sequence of a *Brassica napus* 20-Kilodalton Oleosin Gene¹

Keunmyoung Lee and Anthony H. C. Huang*

Department of Botany and Plant Sciences, University of California, Riverside, California 92521

The storage triacylglycerols in seeds are confined to discrete spherical organelles called oil bodies (10). Each oil body of 1 μ m diameter contains a matrix of triacylglycerols surrounded by a layer of phospholipids embedded with abundant and unique proteins called oleosins (2, 6). Oleosins are hydrophobic proteins of low M_r ranging from 16 to 26 kD, depending on the isoforms and plant species. They are present in the seeds of both dicotyledons and monocotyledons. Their possible functions include stabilizing the oil bodies in an aqueous environment and providing signal recognition for the specific binding of lipase during germination. The partial amino acid sequences of the oleosins from maize (16 kD) (9) and Brassica (20 kD) (4) and the complete amino acid sequences of the oleosins from maize (18 kD) (5) and carrot (19 kD) (1) have been deduced from their cDNA and genomic sequences, respectively.

Here we report the complete nucleotide sequence of a *Brassica napus* oleosin gene (Fig. 1; Table I). The *Brassica* oleosin gene possesses one intron, whereas all the previously reported oleosin genes (two from maize and one from carrot) do not contain introns. The intron in the *Brassica* oleosin gene occurs between the two exons encoding the central hydrophobic domain and the C-terminal domain, respectively.

The deduced amino acid sequence of the *Brassica* oleosin shows the same three basic structural domains that are common to the four sequenced oleosins. These three domains are an N-terminal hydrophilic domain, a central hydrophobic domain, and a C-terminal amphipathic domain. The *Brassica* oleosin shares significant homology in amino acid sequences at the central hydrophobic domain, but not at the other two domains, with the other reported oleosins. The *Brassica* oleosin does not possess appreciable cleavable signal sequence at the N-terminus, and this is in agreement with its M_r of 20,000 estimated by SDS-PAGE being similar to its M_r of 20,682 derived from the amino acid sequence.

The N-terminal amino acid sequence of the *Brassica* oleosin bears little similarity with those of maize 18-kD oleosin and carrot 19-kD oleosin. This is intriguing, because the maize 18-kD oleosin is correctly targeted to the seed oil bodies in *Brassica* transformed with the maize oleosin gene (3). Presumably, the intracellular targeting signal for oleosin resides in

Table I. Characteristics of a 20-kD Oleosin Gene from Brassica	
napus	

Organism:

Brassica napus, L., var Bridger

Location on Chromosome:

Unknown

Function:

Oleosin on the surface of seed oil bodies serving to stabilize the organelles and possibly acting as a receptor for lipase binding during germination.

Gene designation:

BNO20

Source:

Obtained from a λ EMBL-3 genomic library (from Clontech, Palo Alto, CA) using maize 16-kD oleosin cDNA as a probe (9). Fragments subcloned in pUC118 and pUC119 and sequenced by dideoxy sequencing of both strands.

Methods of Identification:

Sequence identical with a segment of the amino acid sequence of an isolated oleosin (8).

Expression and Regulation:

Oleosin synthesized only in maturing seeds. Representing a few percent of the total seed proteins. Oleosin genes in maize (2), carrot (1), and *Brassica* (7) positively regulated by ABA. Putative ABA regulatory elements present at the 5' upstream (Fig. 1).

Features of gene structure:

An open reading frame of 561 bp^2 interrupted by an intron of 462 bp. Intron occurring at the junction of reading frame encoding the central hydrophobic domain and the C-terminal domain. Putative regulatory sequences underlined (Fig. 1).

Codon Usage:

51.3% G + C in the third nucleotide.

(G + C) Content:

51.5% in the coding region. Structural Features of Protein:

ORF of 561 bp encoding a polypeptide of 187 amino acid residues of M_r 20,682, a value similar to the M_r of the isolated oleosin. Three structural domains present in the protein (N-terminal hydrophilic domain, central hydrophobic domain, and C-terminal amphipathic domain). Belonging to the "low M_r " oleosin isoform.

Antibodies:

Prepared earlier (8).

Subcellular Location:

Localized on the surface of oil bodies. EMBL Accession No.

M 63985

¹Support was received from National Science Foundation grant DMB 88–05618.

² Abbreviation: bp, base pair.

-949																														ATGA
-900	CGC	CTT	CCA	GAA	AA	CATC	GAG	CAA	GCI	CTC	GAA	GCA	ACC	AAA	GCT	GAC	CTC	TTT	CGG	ATC	GTA	CAG	AAC	CCG	AAC	AAT	CTC	GTT.	ATG	rccc
-810	GTC	CGT	СТС	CGA	AC	GAC	ATC	CTC	GTA	GCT	CGG	ATT	ATC	GAC	GAA	TCC	ATG	GCT/	ATA	CCC.	AAC	CTC	CGT	CTT	CGT	CAC	GCC	TGG	AAC	CCTC
-720																														FTGC
-630	GCC	GAT.	TGC	CGGC	GGI	AGC	CGG	GTC	GGG	TTG	GGG	ACG	AGA	CCC	GAA	rcco	GAG	CCTC	GGT	GAA	GAG	GTT	GTT	CAT	CGG	AGA	TTT.	ATA	GAC	GGAG
-540	ATC	GGA'	TCG	GAGC	GGI	TTT	GGG	GAA	AGG	GGA	AGT	GGG	TTT	GGC	TCT	TTT	GGA	rag/	AGA	GAG	TGC	AGC	TTT	GGA	GAG	AGA	CTG	GAG	AGG	FTTA
-450	GAG	GAG	AGA		GGG	CGGA	GAI	TAC	CGG	GAGA	GAG	GCG	ACG	AGA	GAT	AGC	ATT	ATCO	GAA	GGG.	AAG	GGA	GAA	AGA	GTG	ACG	TGG.	AGA.	AAT	AGAA
-360	AAC	CCG	TTA	AGA	GTO	GGA	TAT	TTA	TCA	TAT	таа	AAG	ccc	AAT	'GGG	CCTO	GAA	ccci	ATT	ГАА	ACA	AGA	CAG	ATA	AAT	GGG	CCG	TGT	GTT	AAGT
-270																														
-180	ATO	GCG	TGC	ATG	TGZ	AGC	CAI	TAA	CAC	CGTG	GCG	TAG	GAT	GCA	TGA	CGA	CGC	CAT	FGA	CAC	CTG	ACT	CTC	TTC	ССТ	тгт	CTT	CAT	ATA	ICTC
																														CAAA
1	ATG	GAC	GGA	TAC	AGO	TAG	AAC	CCA	TCA		TAT	CAC	AAG	TCG	AGA	FCA	GTA?	rcco	CCG	AGA	CCG	AGA	CCA	GTA	TTC	TAT	GAT	CGG	TCG	AGAC
	M	Т	D	Т	A	R	т	Н	Н	D	I	т	S	R	D	Q	Y	Ρ	R	D	R	D	Q	Y	S	M	I	G	R	D
91	CGA	AGA	CAA	GTA	TTC	CAT	GAT	TGG	CCG	AGG	CCG	AGA	CCA	GTA	CAA	CAT	GTAT	rgg7	rcg/	AGA	CTA	CTC	CAA	GTC	TAG	ACA	GAT	TGC	TAA	GGCT
	R	D	K	Y	S	M	I	G	R	D	R	D	Q	Y	N	M	Y	G	R	D	Y	S	K	S	R	Q	I	A	K	A
181	GTT	FAC	CGC	AGT	CAC	GGC	CGG	TGG	GTC	CCT	тст	TGT	CCT	СТС	CAG	ICT (CAC	CCTT	FGT (CGG.	AAC	FGT	CAT	TGC	TCT	GAC	TGT	TGC	GAC	TCCT
	V	т	A	V	т	A	G	G	S	L	L	v	L	S	S	L	т	L	V	G	T	V	I	A	L	Т	V	A	Т	P
271	CTC	GCT	TGI	TAT	CTT	TAG	TCC	AAT	CCI	TGT	ccc	TGC	TCT	CAI	CAC	CGT	rgc/	ATTO	GCT	CAT	CAC	CGG	CTT	TCT	CTC	CTC	TGG	TGG	CTT	IGGC
	L	L	V	I	F	S	P	I	L	V	P	A	L	Ι	т	V	A	L	L	I	Т	G	F	L	S	S	G	G	F	G
361	2 000			m a m			~		mme		-	~ ` `																	~ ~ ~	acaa
	ALL	reu	RGL	141	AAL	CGT.	UT 1	CTC		GAT	CTA	CAA	gta	agτ	gga	cati	ctaa	aca	τα	τατ	ttc	aag	ttg	tac	aat	atg	τττ	caa	yaa	y vy y
301						V							gta	agτ	gga	cati	ttaa	aaca	ata	τατ	ttc	aag	ttg	tac	aat	atg	TTT	caa	yaa	9099
451	I	A	A	I	т	V	F	S	W	I	Y	K																		
451	I taa	A atti	A tat	I ttt:	T ttt	V ttt:	F ttt	S ttt	W .gaa	I ittt	Y taa	K gaa	ttc	agg	gtt	tcc	ccaa	aago	gct	ttc	tag	gcc	caa	agg	act	ggt	ccc	cct	cct	
451	I taa ctg	A atti jaco	A tat gag	I ttt jctc	T ttt cat	V ttt: gta	F ttt ata	S ttt atg	W gaa ICCC	I ittt :cag	Y taa tgg	K gaa ccc	ttc gag	agg aga	gtt	tcco gtti	ccaa tgca	agg	gct: gtg:	ttc agg	tage ctte	gcc cga	caa	agg	act gcg	ggt tat	ccc tgg	ccto gaa	ccto	ggcg gtaa
451 541 631	I taa ctg ttt	A atti jaco tati	A tat gag taa	I ttt pctc lata	T ttt cat .aaa	V ttt gta cat	F ttt ata aat	S ttt atg ggt	W gaa ccc tga	I ittt cag iata	Y taa tgg tag	K gaa ccc cga	ttc gag cat	agg aga gcc	gtt att	tcc gtti tga	ccaa tgca gggg	agg agcg gaaa	yct ytga aaa	ttc agg aag	tage ctte taca	gcc cga aaa	caa acco cca	agg cgg taa	act gcg aat	ggt tat tat	ccc tgg aca	ccto gaao taao	ccto gcgo ccgo	ggcg gtaa
451 541 631 721	I taa ctg ttt	A atti jaco tati igai	A tat gag taa ttt	I sttt gctc lata stag	T ttt cat aaa ata	V ttt gta icat itta	F ttt ata aat cat	S att atg ggt taa	W gaa ccc tga aat	I ittt cag iata igcc	Y taa tgg tag gtt	K gaa ccc cga tta	ttc gag cat cat	agg aga gcc cat	gtt att ttg	tcco gtti tgao	ccaa tgca gggg	agg agc yaaa tago	gct gtga aaa cta	ttc agg aag tac	tage ctte taca	gcc cga aaa aga	caa acco cca ctto	agg cgg taa gac	act gcg aat tta	ggt tat tat gct	ccc tgg aca agc	ccto gaa taa ttg	ccto gcgo ccga atao	ggcg gtaa acaa cgta
451 541 631 721 811	I taa ctg ttt gtg cgt	A gace tati ggai tgta	A gag taa ttt agt	I sttt sctc lata stag sata	T ttt cat aaa ata tgt	V gta icat itta agc	F ttt ata aat cat atg	S atg ggt taa cac	W gaa ccc tga aat gtg	I scag ata gcc ftgt	Y taa tgg tag gtt gtg	K gaa ccc cga tta aat	ttc gag cat cat tgt	agg aga gcc cat gat	gtt attg ttg cat gaa	tcco gtti tgao ttto tago	ccaa tgca 1999 1990 1990 1990 1990 1990 1990 199	agg agc gaaa tagc rGC A	yct ytga aaa cta AAC T	ttc agg aag tac GGG G	tage ctte taca aca AGA E	gcc cga aaa aga GCA H	caa cca cca ctt CCC	agg cgg taa gac ACA Q	act gcg aat tta AGG G	ggt tat tat gct GTC S	ccc tgg aca agc AGA D	ccto gaa taa ttg: TAA K	cct ccg ata ACT L	ggcg gtaa acaa cgta GGAC D
451 541 631 721	I taa ctg ttt gtg cgt	A gace tati ggai tgta	A gag taa ttt agt	I sttt sctc lata stag sata	T ttt cat aaa ata tgt	V gta icat itta agc	F ttt ata aat cat atg	S atg ggt taa cac	W gaa ccc tga aat gtg	I scag ata gcc ftgt	Y taa tgg tag gtt gtg	K gaa ccc cga tta aat	ttc gag cat cat tgt	agg aga gcc cat gat	gtt attg ttg cat gaa	tcco gtti tgao ttto tago	ccaa tgca 1999 1990 1990 1990 1990 1990 1990 199	agg agc gaaa tagc rGC A	yct ytga aaa cta AAC T	ttc agg aag tac GGG G	tage ctte taca aca AGA E	gcc cga aaa aga GCA H	caa cca cca ctt CCC	agg cgg taa gac ACA Q	act gcg aat tta AGG G	ggt tat tat gct GTC S	ccc tgg aca agc AGA D	ccto gaa taa ttg: TAA K	cct ccg ata ACT L	ggcg gtaa acaa cgta GGAC D
451 541 631 721 811	I taa ctg ttt gtg cgt	A atti gaco cati ggai cgta	A gag taa ttt agt AAG	I sttt gctc lata stag sata	T ttt cat aaa ata tgt GAA	V gta icat itta agc	F ttt ata aat cat atg GGG	S atg atg taa cac AGG	W gaa iccc itga aat igtg	I ittt cag iata gcc itgt	Y taa tgg tag gtt gtg TCA	K gaa ccc cga tta aat GGA	ttc gag cat cat tgt TAT	agg aga gcc cat gat GAA	igtt atto ttg cat gaa .GGA	tcco gtti tgao ttto tago	ccaa tgca gggq ggcl ggcl ggcl ggcl ggcl ggcl ggcl	agg agc gaaa tagc IGCI A ICAC	ycti ytg: aaa; ctai AAC T GTA	ttc agg aag tac GGG G CTA	tage ctte taca acaa AGA E TGG	gcc cga aaa aga GCA H ACA	caa cca cca ctto CCC P ACA	agg cgg taa gac ACA Q GCA	act gcg aat tta AGG G AAC	ggt tat gct GTC S AGG	ccc tgg aca agc AGA D	CCL gaa taa ttg TAA K GGA	CCT GCG CCG ACT L AAC	ggcg gtaa acaa cgta GGAC D GAC
451 541 631 721 811	I taa ctg ttt gtg cgt AGI S	A gaco cati ggai cgta rgci A	A tat gag taa ttt agt AAG R	I sttt jete lata stag ata GAT M	T ttt aaa ata tgt GAJ K	V ttt gta icat itta agc iGCT L	F ttt ata aat cat atg GGG G	S atg ggt taa cac G	W gaa ccc tga aat gtg CAA K	I att cag ata gcc ftgt AGT V	Y taa tgg gtt gtt TCA Q	K gaa ccc cga tta aat GGA D	ttc gag cat cat tgt TAT M	agg aga gcc cat gat GAA K	igtt attg cat igaa GGA	tcco gtti tgao ttto tago CAGi R	ccaa tgca ggci ggci gtai Y AGCI A	agg agc gaaa tago IGC A ICA Q	yct ytg: aaa cta AAC T GTA Y	ttc agg aag tac GGG G CTA Y	tage ctte taca aca AGA E TGG G	gcc cga aga aga GCA H ACA Q	caa acco cca ctto CCC P ACA	agg cgg taa gac ACA Q GCA Q	act gcg aat tta AGG G AAC T	ggt tat gct GTC S AGG G	ccc tgg aca agc AGA D TGG G	cct gaa taa ttg TAA K GGA E	cct gcg ccg ata ACT L AAC	ggcg gtaa acaa cgta GGAC D GAC D
451 541 631 721 811 901	I taa ctg ttt gtg cgt AGI S CGI	A gace gace ggat cgta rgci A rgci	A gag taa ttt agt AAG R CCG	I sttt jete iata stag ata GAT M sTAC	T ttt cat ata tgt GAJ K CCC	V ttt gta icat itta agc iGCT L	F ata aat cat atg GGG G G	S ttt atg ggt taa cac G G CCA	W gaa ccc tga aat gtg CAA K K	I ittt icag iata igcc itgt iAGT V iCAC	Y taa tgg gtt gtt TCA Q TAC	K gaa ccc cga tta aat GGA D CTA	ttc gag cat cat tgt TAT M	agg aga gcc cat gat GAA K	igtt attg cat igaa GGA	tcco gtti tgao ttto tago CAGi R	ccaa tgca ggci ggci gtai Y AGCI A	agg agc gaaa tago IGC A ICA Q	yct ytg: aaa cta AAC T GTA Y	ttc agg aag tac GGG G CTA Y	tage ctte taca aca AGA E TGG G	gcc cga aga aga GCA H ACA Q	caa acco cca ctto CCC P ACA	agg cgg taa gac ACA Q GCA Q	act gcg aat tta AGG G AAC T	ggt tat gct GTC S AGG G	ccc tgg aca agc AGA D TGG G	cct gaa taa ttg TAA K GGA E	cct gcg ccg ata ACT L AAC	ggcg gtaa acaa cgta GGAC D GAC D
451 541 631 721 811 901	I taa ctg ttt gtg cgt AGI S CGI R	A gatti gati ggat cgta IGCI A IGA(D	A gag taa ttt agt AAG R CCG R	I sttt jetc iata stag ata GAT M sTAC T	T ttt cat ata ata tgt GAF K CCC R	V :ttt :gta icat icat icat icat icat icat icat ic	F ata aat cat GGG G G AAC T	S atg ggt taa cac G G CCA Q	W gaa tga tga gtg CAA K GCA H	I ittt cag iata :gcc itgt AGT V ACAC T	Y taa tgg gtt gtt TCA TAC TAC	K gaa ccc cga tta aat GGA D CTA *	ttc gag cat tgt TAT M AAT	agg aga gcc cat gat GAA K TAC	igtt attg cat igaa GGA D GGC	tcco gtti tgao ttgao tago CAGi R ATGi	CCA LgCa JgG1 JgC1 STA1 Y AGC1 A ACT1	agg gaaa tagc IGCI A ICAC Q ATT	gct gtg aaa cta AAC T GTA Y FTC	ttc agg tac GGG CTA Y ATA	tago ctto taca AGA E TGG GTC	gcc cga aga gCA H ACA Q CAA	caa cca cca cct CCC P ACA Q TAA	agg cgg taa gac ACA Q GCA GCA GGC	act gcg aat AGG G AAC T TGA	ggt tat gct GTC S AGG G TGT	CCC tgg aca agc AGA D TGG G CGG	CCt gaa tag TAA GGA GGA GAG	CCT CCG CCG ACT L AAC H ICC	ggcg gtaa acaa cgta GGAC D GAC D AGTT
451 541 631 721 811 901 991	I taa ctg ttt gtg cgt AGI S CGI R TAI	A gatti ggai tgta tgta tgta tGCI A tGA(D tGA(A gag taa ttt agt AAG R CCG R GCA	I Sttt Solata Stag Stag Stac T M STAC T ATA	T ttt aaa ata tgt GAA K CCC R AGG	V Sttt Sgta Scat Sta Sta Sta Sta STGG STGT	F ttt ata cat atg GGG AAC TTA	S ttt atg ggt taa cac AGG G CCA Q AGAA	W gaa ccc tga aat gtg CAA K GCA H TTT	I sttt scag iata gcc itgt AGT V ACAC T TGAT	Y taa tgg gtt gtg TCA TAC TAC	K gaa ccc cga tta aat GGA D CTA TGT	ttc gag cat tgt TAT M AAT	agg aga gcc cat gat GAA K TAC TAA	igtt attg cat igaa GGA D GGC	tcco gtti tgaq tago CAGi R ATGi	CCA LgCa JggG JgCl GTA Y AGCI A ACTA	agg Jaag Jaag Jaag IGCJ A ICA Q ATT GAA	goti gtg: aaa; tai AAC T GTA GTA T GTA GTA	ttc agg tac GGG CTA Y ATA GATA	tago ctto taca aca AGA E TGG GTC ATC	gcc cga aga GCA H ACA Q CAA	caa acco cca ccc p ACA Q TAA	agg cgg taa gac ACA Q GCA GGC	act gcg aat AGG G AAC T TGA	ggt tat gct GTC S AGG G TGT TCT	CCC tgg aca agc AGA D TGG G CGG TTT	CCt gaa taa ttg TAA K GGA GGA GAG	CCTC	ggcg ytaa acaa cgta GGAC D GAC D AGTT ITGT
451 541 631 721 811 901 991	I taa ctg ttt gtg cgt AGI S CGI R TAI TAA	A atti gac cati ggai cgta IGCI A IGAC D IGAC ATT	A tat gag taa ttt agt AAG CCG R CCG R GCA	I Sttt Sctc iata SGAT STAC TAC TATA STGT	T ttt aata ata tgt GAK CCC R AGG	V Sttt Sgta Loat AGCT L STGG TGT TGT	F ttt ata cat GGG GG AAC TTA ATI	S ttt atg ggt taa cac G G G G G A G G A G T G T G T G T G T	W gaa ccc tga aat gtg CAA K GCA H TTT AAA	I sttt scag ata gcc itgt v AGT V CAC T SGAT	Y taa tgg gtt gtg TCA TAC TAC TAC	K gaa ccc cga tta aat GGA D CTA TGT GTA	ttc gag cat tgt TAT M AAT TTA TAT	agg aga gcat gat GAA K TAC TAA	gtt attg cat gaa GGA D GCC TAA	tcco gtti tgao ttto cAGi R ATGi AAGO	CCAR LGCR JGGR JGCT JGCT A A CTI JGGR CTTT	agg Jaa Jaa Jaa TGCJ A TCA Q ATT GAA (CCC)	goti gtg: aaa cta T GTA GTA GTA GTA GTA CTT	ttc agg aag tac 3GG G CTA Y ATA GAT.	tago ctto taca aca AGA E TGG GTC ATC AAG	gcc cga aga GCA H ACA Q CAA ACA IGG	Caa acco cca cct CCC P ACA Q IAA GTC	agg cggg taa gac ACA Q GCA GGC TTT CGT	act gcg aat AGG G AAC T TGA TTT CTA	ggt tat gct GTC AGG TGT TGT TAT	CCC tgg aca agc AGA D TGG G CGG TTT GGT	CCt gaa taa ttg TAA GGA GGA GAG TGG AAA	CCTC ACTO H ICCC ACTO ACCC	ggcg gtaa acaa cgta GGAC D GAC D AGTT FTGT FTAT
451 541 631 721 811 901 991 1081 1171	I taa ctg ttt gtg cgt AGI S CGI R TAI TAA GAI	A Jaco Jaco Jgat Jgat Jgat IGCI A IGAO D IGAO NATI	A tat yag taa ttt agt R CCG R GCA GCA GTC GTC	I sttt ictc iata stag :ata GAT TAC TAC TGT STGT	T ttt cat ata tgt GAM CCC R AGC GTT	V Sttt Sgta Lagc AGCT L STGG G STGT TTTT	F ttt ata cat GGG GG AAC TTA ATT CTC	S atg ggt taa cac AGG G CCA G G AGA TGT TGT	W gaa ccc tga aat CAA K GCA H TTTI AAA	I ittt icag iata igcc itgt iAGT V iCAC T iCAC T iCAC	Y taa tgg gt gt TCA TAC TAC TAC TAA	K gaa ccc cga tta aat GGA D CTA * TGT GTA AA	ttc gag cat tgt TAT M AAT TTA TAT GAC	agg aga gcat gat GAA K TAC TAC TGC	GGA GGA GGA GGC GCC TAA GTA	tcco gtti tgaq tago CAGi R ATGi AAGO CTTO	CCA tgCa JgGC JGC JGC JGC JGC A A A CT J A CT J A CT J TATO	agg Jaag Jaag TGCA PCA QATT GAA CCC CTT	get gtg aaa ta T GTA GTA GTA GTA GAT CTT TAG	ttc agg tac GG CTA CTA Y ATA GAT. FTTA	tage ctt taca aca AGA E TGG G G TC AAG TAT	gcca cga aga GCA H ACA Q CAA CAA TGG TAT	Caa cca cca ccc P ACA Q TAA GTC GTT	agg cgg taa gac ACA Q GCA GGC ITT GAG	act gcg aat AGG G AAC T TGA TTT TTT TAA	ggt tat gct GTC S GTC TGT TGT TCT TATG	CCC tgg aca agc AGA TGG G CGG TTT GGT AAC	CCT gaa taa ttg TAA GGA E GAG TGG AAA TTT	CCT ACT ACT ACT ACC TCC CCT CCT CCT	ggcg gtaa acaa cgta GGAC D GAC D AGTT ITGT ITGT ITAT AGAT
451 541 631 721 811 901 991 1081 1171 1261	I taa ctg ttt gtg cgt AGI S CGI R TAI TAA GAI CTG	A Jacq Jacq Jgat Jgat Jgat IGCI A IGA IGA J IGA J JTG GGT	A tat gag taa ttt agt R CCG R CCG CCG GCA GCA GCA	I ttt icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta icta	T tttt cat aaa ata tgt GAM K CCCC R AGG GTI CCCI GAG	V Sttt Sgta L Sta L STGC STGG STGT STGT STAG	F ttta aat GGG G AAC TTA ATI CTC ACI	S atg ggt taa gggt taa GAGG GAA TGT TGT TGT	W gaa ccc tga aat CAA K GCA H TTA TTA AGC	I ittt icag igcc igcc igcc igcc icgc I icac T icac I icag icgc I icag icgc I icag I icag I icag I icag I I I I I I I I I I I I I	Y taa tgg gtg TCA TAC TAC TAC TAA TAA GTT	K gaa ccc cga tta aat GGA TGT GTA GAG	ttc gag cat tgt TAT M AAT TAT GAC CTG	agg aga gcc cat gat GAA K TAC TAC GTI TGC AAC	GCC. GCC. GCC. GCC. GCC. GCC. GCC. GCC.	tcco gtti tgaq ttto tago CAGi R ATGi ATGi ATGI	CCAR LGCR JGGS JGCI JGCI AGCI ACTI ACTI SGGG CTTI FATC GCTC	Aagg Jaaa Lago IGCA Q ATTT GAAC CTTT GGCA	goti Jtg aaa Sta AAC T GTA T GTA GTA CTT TAG	ttc: agg tac; GGG CTA Y ATA GAT. FTT. GTG	tage ctt taca aca AGA G G G G G C A A A A C A A A C	gcca cga aga GCA H ACA Q CAA CAA TGG TAT ACT	Caa cca cca cct CCC P ACA Q IAA GTC GTC GGA	agg cgg taa gac ACA Q GCA GGC FTT CGT GAG IGC	act gcg aat AGG G AAC TGA TGA TTT CTA TAA	ggt tat gct S GTC S G TGT TGT TAT ATG	CCCC aca agc AGA TGG CGG TTT GGT AAC AGA	cct gaa taa ttg TAA GGA GGA GAG TGG TGG	CCT ACT ACT ACT ACC TCC ACC CCT GAA	ggcg gtaa acaa ggac D JAC D AGTT TTGT TTAT AGAT GATC
451 541 631 721 811 901 991 1081 1171 1261 1351	I taa ctg ttt gtg cgt AGI S CGI R TAI TAA GAI CTG TCI	A atti jacq cati cgta cgta rGCI D rGA(AAT) rGA(GGT) rAA)	A tat gag taa ttt AAG CCG CCG GCA TTG CCG TAT	I Sttt Stata Stata STAC TTT STAC STAC STAC STAC STAC STAC	T tttt cat aaa ata tgt GAM K CCCC R AGC GTT CCCT GAC	V Sttt Sgta L Stat Cat SGG TGT STGG STGT STGG STGT STAG SGGG	F tttt ata aat GGG GGG TTA AAC TTA ATTI CTCC ACT	S ttt atg ggt taa GAG GAA TGT TGT AGC GAA	W gaa ccc tga aat gtg CAA K GCAA H TTT TAAA CAA CAA	I attt cag ata gcc itgt AGT V CAC T CGAT CGA CGA	Y taa tgg gtt gt TCA TAC TAC TAA TAA TAA GTT GTG	K gaa ccga tta aat GGA CTA TGTA GAG GAG TCT	ttc gag cat tgt TAT AAT TTA GAC CTG ATA	agg aga gcc cat gat TAC GTI TAC GTI TAC TTI	GTA GGA GGA GGA GGA GCC GTA GTA GTA GTA TGA	tccc tgtt tgaq tag CAG R ATG ATG TTT ACTO	CCA LGC2 JGC1 JGC1 JGC1 JGC1 A ACTI A CTT FATC GCTC GCA	ago Jaaa Lago IGCI A ICAO Q ATTI GAAO ICCO CTTI GGCI ITAI	JCt: Jtg: Aaa: Cta: T GTA: Y FTC: GAT(CTT) AAT(AGC)	ttc agg tac GGG G CTA Y ATA GAT. GTG GTG	tage cttc aca: AGA E TGG GTC ATC AAG TAT AAC	JCC Cga aga GCA H ACA Q CAA CAA TGG TAC TAC TAC	Caa acca cct CCC P ACA Q TAA GTC GTC GGA AGA	agg taa gac Q GCA Q GGC TTT CGT GAG GAG TGC	act gcg aat AGG AAC TGA TGA TTT TAA AAG AAC	ggt tat gctC GTC GTC TGT TAT ATC ATC TGA	CCC tgg aca agc AGA TGG CGG TTT GGT AAC AGA TGC	CCT gaa taa TAA GGA GGA GAG TGG TGG TGT AGT	CCTC CCTC ACTO ACTO H FCCA CCTT CATC CATC	ggcg gtaa cgta GGAC D SAC D AGTT ITGT ITGT ITAT AGAT GATC IGGT

Figure 1. A 2640-bp² sequence of a *B. napus* 20-kD oleosin gene. The transcription initiation codon ATG is numbered as the first three nucleotides. The putative regulatory sequences are underlined, including TATA box, CAAT boxes, RY repeats, GC boxes, octamer, putative ABA-binding sites (ACGTGGCGT and ACGTGTCCT), and polyadenylation signal. The intron sequence is printed in lower case letters. The predicted amino acid sequence is shown. The amino acid sequence, which was determined earlier by peptide microsequencing (8), is underlined.

the highly conserved central hydrophobic domain or in the yet-unidentified secondary structures in the N- or C-terminus.

Earlier, we identified at least two isoforms of oleosins in seeds from different, and within the same, species (8). The two isoforms are distinguishable immunologically, and in general, one isoform has a M_r higher than that of the other isoform within the same species. The Brassica 20-kD oleosin and the maize 16-kD oleosin apparently belong to the same "low M_r " isoform, whereas the carrot 19-kD and maize 18kD represent the "high M_r " isoform. This category is revealed in a comparison of the amino acid sequences among the oleosins: 48% identical amino acid residues between the two high Mr oleosins (maize 18-kD oleosin and carrot 19-kD oleosin), 51% between the two low M_r oleosins (maize 16-kD oleosin and rapeseed 20-kD oleosin), 37% between the two maize oleosins, and 33% or less between any other pair of high and low M_r oleosins. The findings that each pair of oleosin isoforms are more related between themselves than to the other oleosin isoforms explains our earlier observation of their immunological kinship (8).

LITERATURE CITED

- Hatzopoulos P, Franz G, Choy L, Sung RZ (1990) Interaction of nuclear factors with upstream sequences of a lipid body membrane protein gene from carrot. Plant Cell 2: 457–467
- Huang AHC, Qu R, Lai YK, Ratnayake C, Chan KL, Kuroki GW, Oo KC, Cao YZ (1991) Structure, synthesis and degradation of oil bodies in maize. *In* MJ Emes, ed, Compartmentation of Plant Metabolism in Non-Photosynthetic Tissues. Cambridge University Press, Cambridge, England, pp 43-58
- 3. Lee WS, Tzen JTC, Kridl JC, Radke SE, Huang AHC (1991) Maize oleosin is correctly targeted to seed oil bodies in *Brassica napus* transformed with the maize oleosin gene. Proc Natl Acad Sci USA in press
- Murphy DJ, Keen JN, O'Sullivan JN, Au DMY, Edwards EW, Jackson PJ, Cummins I, Gibbons T, Shaw CH, Ryan AJ (1991) A class of amphipathic proteins associated with lipid storage bodies in plants. Possible similarities with animal serum apolipoproteins. Biochim Biophys Acta 1088: 86–94
- Qu R, Huang AHC (1990) Oleosin KD18 on the surface of oil bodies in maize: genomic and cDNA sequences, and the deduced protein structure. J Biol Chem 265: 2238-2243

- Qu R, Wang SM, Lin YH, Vance VB, Huang AHC (1986) Characteristics and biosynthesis of membrane proteins of lipid bodies in the scutella of maize (*Zea mays L.*). Biochem J 234: 57-65
- Taylor DC, Weber N, Underhill EW, Pomeroy MK, Keller WA, Scowcroft WR, Wilen RW, Moloney MM, Holbrook LA (1990) Storage protein regulation and lipid accumulation in microspore embryos of *Brassica napus* L. Planta 181: 18-26
- Tzen JTC, Lai YK, Chan KL, Huang AHC (1990) Oleosin isoforms of high and low molecular weights are present in the oil bodies of diverse seed species. Plant Physiol 94: 1282-1289
- Vance VB, Huang AHC (1987) The major protein from lipid bodies of maize. Characteristics and structure based on cDNA cloning. J Biol Chem 262: 11275-11279
- Yatsu LY, Jacks TJ (1972) Spherosome membranes. Half unitmembrane. Plant Physiol 49: 937–943