Genome-Wide Analysis of Oleosin Gene Family in 22 Tree Species: An Accelerator for Metabolic Engineering of BioFuel Crops and Agrigenomics Industrial Applications?

Heping Cao

Abstract

Trees contribute to enormous plant oil reserves because many trees contain 50%–80% of oil (triacylglycerols, TAGs) in the fruits and kernels. TAGs accumulate in subcellular structures called oil bodies/droplets, in which TAGs are covered by low-molecular-mass hydrophobic proteins called oleosins (OLEs). The OLEs/TAGs ratio determines the size and shape of intracellular oil bodies. There is a lack of comprehensive sequence analysis and structural information of OLEs among diverse trees. The objectives of this study were to identify OLEs from 22 tree species (e.g., tung tree, tea-oil tree, castor bean), perform genome-wide analysis of OLEs, classify OLEs, identify conserved sequence motifs and amino acid residues, and predict secondary and three-dimensional structures in tree OLEs and OLE subfamilies. Data mining identified 65 OLEs with perfect conservation of the “proline knot” motif (PX5SPX3P) from 19 trees. These OLEs contained >40% hydrophobic amino acid residues. They displayed similar properties and amino acid composition. Genome-wide phylogenetic analysis and multiple sequence alignment demonstrated that these proteins could be classified into five OLE subfamilies. There were distinct patterns of sequence conservation among the OLE subfamilies and within individual tree species. Computational modeling indicated that OLEs were composed of at least three α -helices connected with short coils without any β -strand and that they exhibited distinct 3D structures and ligand binding sites. These analyses provide fundamental information in the similarity and specificity of diverse OLE isoforms within the same subfamily and among the different species, which should facilitate studying the structure-function relationship and identify critical amino acid residues in OLEs for metabolic engineering of tree TAGs.

Introduction

TRIACYLGLYCEROLS (TAGs) ARE THE MAJOR MOLECULES of energy storage in eukaryotes, serve as a reservoir of fatty acids for membrane biogenesis, and lead to obesity due to excessive accumulation in adipose tissues (Chapman et al., 2012; Dyer et al., 2008; Farese Jr. and Walther, 2009; Somerville et al., 2000). Plant TAGs are rich source of edible oils for human consumption and special oils for industrial applications (Cahoon et al., 2007; Dyer et al., 2008) and biofuels (Carvalho et al., 2008; Chen et al., 2010b; Misra et al., 2013; Tai and Stephanopoulos, 2013; Winichayakul et al., 2013). Trees contribute to enormous plant oil reserves because fruits and kernels of many trees contain up to 50%–80% of oil (Cao et al., 2013; Chen et al., 2008; Zeng et al., 2014; Zhang et al., 2014).

TAGs accumulate in discrete subcellular organelles called oil bodies in plants (Huang, 1992), similar to oil droplets in

animals (Walther and Farese Jr, 2012). Plant oil bodies mainly consist of TAGs surrounded by a monolayer of phospholipids, with the hydrophobic acyl moieties of the phospholipids interacting with TAGs and the hydrophilic head groups facing the cytosol (Chapman et al., 2012; Huang, 1996). In addition, plant oil bodies contain a number of proteins including oleosins and caleosins (Jolivet et al., 2013).

Oleosins (OLEs) are a group of mostly low-molecular mass (15–25 kDa) hydrophobic proteins localized on the surfaces of plant oil bodies (Dyer et al., 2008; Huang, 1992). The precise functions of OLEs are to be determined. They may stabilize oil bodies at low water potential and/or regulate the size of oil bodies (Somerville et al., 2000). The OLEs/TAGs ratio determines the size and shape of intracellular oil bodies (Ting et al., 1996). The relative net amounts of OLE and TAG accumulated during seed development are a major factor determining oil-body size in desiccation-tolerant seeds

(Sarmiento et al., 1997). Changes of oil body sizes cause disruption of storage organelles, alter accumulation of lipids and proteins, and delay seed germination (Siloto et al., 2006). Large oil bodies are highly correlated with lower oil content (Hu et al., 2009). OLEs increase the viability of over-wintering oilseeds by preventing abnormal fusion of oil bodies during imbibitions in the spring (Shimada et al., 2008).

OLE is proposed to be a surfactant because knockdown of the 24-kDa OLE produces micro-oil bodies that aggregate into oil body/ER complexes in soybean (Schmidt and Herman, 2008). Tapetal lipid droplets are required for the biogenesis of the pollen coat of *Brassica napus* (Hernandez-Pinzon et al., 1999). Rapid initiation of *Arabidopsis* pollination requires the oleosin-domain protein GRP17 (Mayfield and Preuss, 2000). Recent research suggests that OLEs may be bifunctional enzymes with both monoacylglycerol acyltransferase and phospholipase activities regulated by serine/threonine/tyrosine protein kinases (Parthibane et al., 2012a; 2012b).

The first OLE gene in plants was identified from *Zea mays* as a partial clone in 1987 and the full-length sequence was determined in 1990 (Qu and Huang, 1990; Vance and Huang, 1987). More OLE genes were isolated from soybean (Kalinski et al., 1991), sunflower (Cummins and Murphy, 1992), and *Brassica napus* (Keddie et al., 1992a; 1992b). The first *Arabidopsis* OLE sequence reported was isolated from its genomic library using *Brassica napus* oleosin probe (Rooijen et al., 1992). The complete genome of *Arabidopsis thaliana* has been sequenced in 2000 (Lin et al., 1999; Mayer et al., 1999; Salanoubat et al., 2000; Tabata et al., 2000; Theologis et al., 2000).

The first tree OLEs were reported in 1995 in *Citrus sinensis* (sweet orange) (Naot et al., 1995) and *Prunus dulcis* (almond) (Garcia-Mas et al., 1995). Since then, various forms of tree OLEs had been identified in *Elaeis guineensis* (oil palm) (Yong et al., 1999), *Theobroma cacao* (cacao) (Guiloteau et al., 2003), *Ricinus communis* (castor bean) (Eastmond, 2004), *Prunus armeniaca* (apricot) (Geuna et al., 2005), *Coffea arabica* (coffee) (Simkin et al., 2006), *Coffea canephora* (coffee) (Simkin et al., 2006), *Corylus avellana* (hazelnut) (Akkerdaas et al., 2006), *Camellia oleifera* (tea oil), *Olea europaea* (olive) (Giannoulia et al., 2007), *Persea americana* (avocado) (Lopez-Gomez et al., 2007), *Cocos nucifera* (coconut palm) (Regalado et al., 2008), *Ficus pumila* (climbing fig) (Chua et al., 2008), *Cupressus sempervirens* (pencil pine) (Pedron et al., 2009), *Pinus taeda* (loblolly pine) (Kovach et al., 2010), *Vernicia fordii* (tung tree) (Cao et al., 2014; Chen et al., 2010a; Long et al., 2010), *Jatropha curcas* (barbados nut) (Gu et al., 2012; Popluechai et al., 2011; Xu et al., 2011), and *Juglans regia* (walnut).

Most importantly, the sequences of five tree genomes have been reported, including *Populus trichocarpa* (poplar) (Tuskan et al., 2006), *Vitis vinifera* (grapevine) (Jaillon et al., 2007), *Ricinus communis* (castor bean) (Chan et al., 2010), *Prunus persica* (peach) (Verde et al., 2013), and *Theobroma cacao* (cacao) (Motamayor et al., 2013).

Plant omics and systems science research is at the epicenter of 21st century integrative biology (Sahu et al., 2014). The immediate challenge of post-genomic biology is to determine the biological functions of proteins encoded by unknown genes. Approaches to gain clues about the structure–function relationship of proteins are to perform comprehensive amino acid sequence analysis (Cao, 2011), predict

their structures at the secondary and three-dimensional levels (Hyun et al., 2013; Misra and Panda, 2013; Pastor et al., 2013), and solve the structures of important proteins (Wang et al., 2001; 2002; Wang and Tanaka Hall, 2001).

It is generally accepted that critical amino acid residues and sequence motifs in the same family of proteins are evolutionarily conserved (Beatty et al., 1999; Cao, 2011; Cao et al., 1999). There is a lack of comprehensive analysis of amino acid sequences and structural information of OLEs among diverse trees. The objectives of this study were to identify OLEs from tree species, perform genome-wide analysis of tree OLEs, classify tree OLEs, identify conserved sequence motifs and amino acid residues, and predict secondary and three-dimensional structures in all tree OLEs and OLE subfamilies for better understanding of the structure–function relationship of these important proteins.

Methods

Database search of OLEs

OLE sequences were obtained from keyword search with “oleosin” and Blastp search (Altschul et al., 1990; Altschul and Lipman, 1990) using tung tree (*Vernicia fordii*) OLE (VfoOLE) sequences (Cao et al., 2014) against the National Center for Biotechnology Information (NCBI)’s non-redundant protein sequence databases (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The keyword search should pick up any sequence entry with the word “oleosin.” Using tung tree oleosin sequences for Blastp search was an additional measure to confirm that all relevant oleosin sequences from tree species were identified from the database. A total of 81 OLEs were obtained from 22 trees. The names of OLEs used in the analysis and their corresponding organisms, classification of OLE subfamily, and the GenBank accession numbers are presented in Tables 1 and 2. The name of each protein sequence consists of the initials of the organism followed by OLE and the GenBank accession number.

Protein analysis

The properties and amino acid compositions of OLEs were analyzed using Vector NTI software (Invitrogen) (Zimmerman et al., 1968). Statistics was performed using Microsoft Excel (Cao, 2011).

Phylogenetic analysis

Phylogenetic analysis was used to study the presumed evolutionary relationships among all OLEs. This analysis was performed using the Vector NTI software (Invitrogen) based on the Neighbor-Joining method of Saitou and Nei (1987). The numbers in the parenthesis following OLE names are the calculated distance values which reflect the degree of divergence between all pairs of OLE sequences analyzed.

Multiple sequence alignment

Multiple sequence alignment was performed using the ClustalW algorithm (Larkin et al., 2007; Thompson et al., 1994) of the AlignX program of the Vector NTI software. This method is based on algorithms that assign scores to aligned residues and detect sequence similarities. Identical

TABLE 1. 46 OLE SEQUENCES FROM 5 COMPLETED TREE GENOMES*

Tree	Organism	Abbreviation	Ole No.	Ole name	GenBank #	Locus (Chromosome)	Proline Knot	Reference
1	<i>Prunus persica</i> (peach)	Ppe	1	PpeOle-EMI10773.1	EMI10773.1	PRUPE_ppa011111mg (NA)	PX5SPX3P	(Verde et al. 2013)
			2	PpeOle-EMI10882.1	EMI10882.1	PRUPE_ppa012323mg (NA)	PX5SPX3P	(Verde et al. 2013)
			3	PpeOle-EMI16302.1	EMI16302.1	PRUPE_ppa012747mg (NA)	PX5SPX3P	(Verde et al. 2013)
			4	PpeOle-EMI19764.1	EMI19764.1	PRUPE_ppa012948mg (NA)	PX5SPX3P	(Verde et al. 2013)
			5	PpeOle-EMI25876.1	EMI25876.1	PRUPE_ppa025868mg (NA)	PX5SPX3P	(Verde et al. 2013)
			6	PpeOle-EMI26259.1	EMI26259.1	PRUPE_ppa018784mg (NA)	PX5SPX3P	(Verde et al. 2013)
2	<i>Populus trichocarpa</i> (poplar)	Ptr	1	PtrOle ¹ -XP_002322186.1	XP_002322186.1	POPTRDRAFT_252246 (LGXV)	PX5SPX3P	(Tuskan et al. 2006)
			2	PtrOle-XP_002297927.1	XP_002297927.1	POPTRDRAFT_708981 (LGI)	PX5SPX3P	(Tuskan et al. 2006)
			3	PtrOle-XP_002300049.1	XP_002300049.1	POPTRDRAFT_641631 (LGI)	PX5SPX3P	(Tuskan et al. 2006)
			4	PtrOle-XP_002304591.1	XP_002304591.1	POPTRDRAFT_646862 (LGIII)	PX5SPX3P	(Tuskan et al. 2006)
			5	PtrOle-XP_002309527.1	XP_002309527.1	POPTRDRAFT_653963 (LGI)	PX5SPX3P	(Tuskan et al. 2006)
			6	PtrOle-XP_002317928.1	XP_002317928.1	POPTRDRAFT_569851 (LGXII)	PX5SPX3P	(Tuskan et al. 2006)
			7	PtrOle-XP_002318659.1	XP_002318659.1	POPTRDRAFT_422500 (LGXII)	PX5SPX3P	(Tuskan et al. 2006)
			8	PtrOle-XP_002323804.1	XP_002323804.1	POPTRDRAFT_736170 (LGXVII)	PX5SPX3P	(Tuskan et al. 2006)
			9	PtrOle-XP_002334246.1	XP_002334246.1	POPTRDRAFT_670493 (NA)	PX5SPX3P	(Tuskan et al. 2006)
3	<i>Ricinus communis</i> (castor bean)	Rco	1	RcoOle-XP_002511014.1	XP_002511014.1	RCOM_1502140A (NA)	PX5SPX3P	(Chan et al. 2010)
			2	RcoOle-XP_002511342.1	XP_002511342.1	RCOM_1509400A (NA)	PX5SPX3P	(Chan et al. 2010)
			3	RcoOle-XP_002511984.1	XP_002511984.1	RCOM_1618240A (NA)	PX5SPX3P	(Chan et al. 2010)
			4	RcoOle-XP_002516493.1	XP_002516493.1	RCOM_0799030A (NA)	PX5G ² PX3P	(Chan et al. 2010)
			5	RcoOle-XP_002521458.1	XP_002521458.1	RCOM_1064090A (NA)	PX5G ² PX3P	(Chan et al. 2010)
			6	RcoOle-AAR15171.1	AAR15171.1	AAR15171 (=RcoOle ⁵)	PX5SPX3P	(Eastmond 2004)
			7	RcoOle-AAR15172.1	AAR15172.1	AAR15172 (=RcoOle ¹)	PX5SPX3P	(Eastmond 2004)
4	<i>Theobroma cacao</i> (cacao)	Tca	1	TcaOle-EOY02487.1	EOY0248.1	TCM_016949 (4)	PX5SPX3P	(Motamayor et al. 2013)
			2	TcaOle-EOY11882.1	EOY11882.1	EOY11882 (7)	PX5SPX3C	(Motamayor et al. 2013)
			3	TcaOle-EOY20852.1	EOY20852.1	EOY20852 (3)	PX5SPX3P	(Motamayor et al. 2013)
			4	TcaOle-EOY20922.1	EOY20922.1	EOY20922 (3)	PX5SPX3P	(Motamayor et al. 2013)
			5	TcaOle-EOY22772.1	EOY22772.1	TCM_014844 (3)	PX5SPX3P	(Motamayor et al. 2013)
			6	TcaOle-EOY29822.1	EOY29822.1	TCM_037237 (9)	PX5SPX3P	(Motamayor et al. 2013)
			7	TcaOle-AAM46777.1	AAM46777.1	AF466102_1 (=TcaOle ¹)	PX5SPX3P	(Guilloteau et al. 200)
			8	TcaOle-AAM46778.1	AAM46778.1	AF466103_1 (=TcaOle ⁶)	PX5SPX3P	(Guilloteau et al. 200)
5	<i>Vitis vinifera</i> (grapevine)	Vvi	1	VviOle-CBII5408.3	CBII5408.3	VIT_00008228001 (NA)	PX5SPX3P	(Jailion et al. 2007)
			2	VviOle-XP_002273158.2	XP_002273158.2	LOC100255962 (14)	PX5SPX3P	(Jailion et al. 2007)
			3	VviOle-XP_002273242.1	XP_002273242.1	LOC100244020 (17)	PX5SPX3P	(Jailion et al. 2007)
			4	VviOle-XP_002275021.1	XP_002275021.1	LOC100252638 (14)	PX5SPX3P	(Jailion et al. 2007)
			5	VviOle-XP_002275087.1	XP_002275087.1	LOC100250821 (14)	PX5SPX3P	(Jailion et al. 2007)
			6	VviOle-XP_002275496.1	XP_002275496.1	LOC100262917 (4)	PX5SPX3P	(Jailion et al. 2007)
			7	VviOle-XP_002281381.1	XP_002281381.1	LOC100256947 (17)	PX5SPX3P	(Jailion et al. 2007)
			8	VviOle-XP_002281746.1	XP_002281746.1	LOC100249321 (16)	PX5SPX3P	(Jailion et al. 2007)
			9	VviOle-XP_003631370.1	XP_003631370.1	LOC100854621 (2)	PX5SPX3P	(Jailion et al. 2007)
			10	VviOle-XP_003634256.1	XP_003634256.1	LOC100256947 (=VviOle ⁷)	PX5SPX3P	(Jailion et al. 2007)
			11	VviOle-CAN68989.1	CAN68989.1	VITISV_017729 (=VviOle ⁹)	PX5SPX3P	(Velasco et al. 2007)
			12	VviOle-CAN69350.1	CAN69350.1	VITISV_012978 (=VviOle ⁸)	PX5SPX3P	(Velasco et al. 2007)
			13	VviOle-CAN74835.1	CAN74835.1	VITISV_023324 (=VviOle ⁶)	PX5SPX3P	(Velasco et al. 2007)
			14	VviOle-CAN80216.1	CAN80216.1	VITISV_043128 (=VviOle ²)	PX5SPX3P	(Velasco et al. 2007)
			15	VviOle-CAN80218.1	CAN80218.1	VITISV_043130 (=VviOle ⁵)	PX5SPX3P	(Velasco et al. 2007)
			16	VviOle-CAN80922.1	CAN80922.1	VITISV_013175 (=VviOle ⁷)	PX5SPX3P	(Velasco et al. 2007)

35 unique OLE, 32 OLE with perfect 'proline knot' motif, 31 full-length OLEs.

Colored amino acid residues in "Proline Knot" column were diverged from the consensus motif sequences. Highlighted names in "Locus" column represent duplication or incomplete sequences.

TABLE 2. 48 OLE SEQUENCE FROM 17 OTHER TREE SPECIES*

Tree	Organism	Abbreviation	Ole No.	Ole name	GenBank #	Coded by	Proline Knot	Reference
6	<i>Coffea arabica</i> (coffee)	Car	1	CarOle-AAAY14574.1	AAY14574.1	AY928084.1:1..447	PX5SPX3P	(Simkin et al. 2006)
7	<i>Coffea canephora</i> (coffee)	Cca	1	CcaOle-AAAX49389.1	AAX49389.1	AY841271.1:79..525	PX5SPX3P	(Simkin et al. 2006)
			2	CcaOle-AAAX49390.1	AAX49390.1	AY841272.1:87..494	PX5SPX3P	(Simkin et al. 2006)
			3	CcaOle-AAAX49391.1	AAX49391.1	AY841273.1:66..593	PX5SPX3P	(Simkin et al. 2006)
			4	CcaOle-AAAX49392.1	AAX49392.1	AY841274.1:74..502	PX5SPX3P	(Simkin et al. 2006)
			5	CcaOle-AAAX49393.1	AAX49393.1	AY841275.1:32..535	PX5SPX3P	(Simkin et al. 2006)
8	<i>Corylus avellana</i> (hazelnut)	Cav	1	CavOle-AAO65960.1	AAO65960.1	AY224599.1:1..423	PX5SPX3P	(Akkerdaas et al. 2006)
			2	CavOle-AAO67349.2	AAO67349.2	AY224679.2:1..480	PX5SPX3P	(Akkerdaas et al. 2006)
9	<i>Cocos nucifera</i> (coconut palm)	Cnu	1	CnuOlep-ACH91011.1	ACH91011.1	FJ012155.1: <1..176	X5SPX3P	(Regalado et al. 2008)
			2	CnuOlep-ACH91012.1	ACH91012.1	FJ012156.1: <1..242	X5SPX3P	(Regalado et al. 2008)
			3	CnuOlep-ACH91013.1	ACH91013.1	FJ012157.1: <1..233	X5SPX3P	(Regalado et al. 2008)
10	<i>Camellia oleifera</i> (tea oil)	Col	1	ColOle-ABF57559.1	ABF57559.1	DQ489495.1:45..491	PX5SPX3P	NP
			2	ColOle-ABF57560.1	ABF57560.1	DQ489496.1:72..497	PX5SPX3P	NP
			3	ColOle-ABF57561.1	ABF57561.1	DQ489497.1:72..536	PX5SPX3P	NP
			4	ColOle-ABF57562.1	ABF57562.1	DQ489498.1:68..541	PX5SPX3P	NP
			5	ColOle-ABF57563.1	ABF57563.1	DQ489499.1:22..444	PX5SPX3P	NP
11	<i>Cupressus sempervirens</i> (pencil pine)	Cse	1	CseOlep-ACA30297.1	ACA30297.1	EU430721.1: <1..>392	X5SPX3P	(Pedron et al. 2009)
			2	CseOlep-ACI87763.1	ACI87763.1	FJ237443.1: <1..268	X5SPX3P	(Pedron et al. 2009)
			3	CseOlep-ACI87801.1	ACI87801.1	FJ237481.1: <1..>389	X5SPX3P	(Pedron et al. 2009)
			4	CseOlep-ACI87814.1	ACI87814.1	FJ237494.1: <1..>293	TX5SPX3T	(Pedron et al. 2009)
			5	CseOlep-ACI09660.1	ACI09660.1	FJ380021.1: <1..>376	TX5SPX3T	(Pedron et al. 2009)
12	<i>Citrus sinensis</i> (sweet orange)	Csi	1	CsiOle-CAA88360.1	CAA88360.1	NA	PX5SPX3P	(Naot et al. 1995)
13	<i>Elaeis guineensis</i> (African oil palm)	Egu	1	EguOlep-AAD41080.1	AAD41080.1	AF147758.1:2..253	MX5SPX3P	(Yong et al. 1999)
			2	EguOle-AF273023_1	AAF76238.1	AF273023.1: <1..424	PX5SPX3P	NP
14	<i>Ficus pumila</i> (climbing fig)	Fpu	1	FpuOle-ABQ57396.1	ABQ57396.1	EF583922.1:29..496	PX5SPX3P	(Chua et al. 2008)
			2	FpuOle-ABQ57397.1	ABQ57397.1	EF583923.1:11..472	PX5SPX3P	(Chua et al. 2008)
15	<i>Jatropha curcas</i> (barbados nut)	Jcu	1	JcuOle-ABW90148.2	ABW90148.2	EU234462.2:68..511	PX5SPX3P	(Popluechai et al. 2011)
			2	JcuOle-ABW90149.2	ABW90149.2	EU234463.2:51..518	PX5SPX3P	(Popluechai et al. 2011)
			3	JcuOle-ABW90150.2	ABW90150.2	EU234464.2:73..486	PX5SPX3P	(Popluechai et al. 2011)
			4	JcuOle-AFP19884.1	AFPI9884.1	JX073622.1:25..492 (= JcuOle2)	PX5SPX3P	(Xu et al. 2011)
			5	JcuOle-AFP19885.1	AFPI9885.1	JX073623.1:53..466	PX5SPX3P	(Gu et al. 2012)
16	<i>Juglans regia</i> (English walnut)	Jre	1	JreOle-AET74076.1	AET74076.1	JN409447.1:1..423	PX5SPX3P	NP
			2	JreOle-AET74077.1	AET74077.1	JN409448.1:1..420	PX5SPX3P	NP
17	<i>Olea europaea</i> (olive)	Oeu	1	OeuOle-AAL92479.1	AAL92479.1	AY083161.1:47..544	PX5SPX3P	(Giannoulia et al. 2007)
18	<i>Persea americana</i> (avocado)	Pam	1	PamOle-AGT63296.1	AGT63296.1	KF006324.1:1..480	PX5SPX3P	(Lopez-Gomez et al. 2007)
19	<i>Prunus armeniaca</i> (apricot)	Par	1	ParOlep-AAAY34921.1	AAAY34921.1	AY962832.1: <1..>87 (29aa)	PX5SPX3P	(Geuna et al. 2005)
20	<i>Prunus dulcis</i> (almond)	Pdu	1	PduOle-Q43804.1	Q43804.1	X78118.1:47..493	PX5SPX3P	(Garcia-Mas et al. 1995)
			2	PduOle-CAA55008.1	CAA55008.1	FB868920.1:1..447 (= PduOle1)	PX5SPX3P	NP
			3	PduOle-CAW72962.1	CAW72962.1	HB738979.1:1..447 (= PduOle1)	PX5SPX3P	NP
			4	PduOle-CBC64401.1	CBC64401.1	HB802593.1:1..447 (= PduOle1)	PX5SPX3P	NP
			5	PduOle-CBD18275.1	CBD18275.1	X78118.1:47..493 (= PduOle1)	PX5SPX3P	(Garcia-Mas et al. 1995)
21	<i>Pinus taeda</i> (loblolly pine)	Pta	1	PtaOlep-AFG55850.1	AFG55850.1	FJ094804.2: <1..309	PX5SPX3P	(Kovach et al. 2010)
			2	PtaOlep-AFG55851.1	AFG55851.1	FJ094810.2: <1..309	PX5SPX3P	(Kovach et al. 2010)
22	<i>Vernicia fordii</i> (tung tree)	Vfo	1	VfoOle-GU245884	GU245884	NA	PX5SPX3P	(Long et al. 2010)
			2	VfoOle-GU245885	GU245885	NA	PX5SPX3P	(Long et al. 2010)
			3	VfoOle-GR217754	GR217754	NA	PX5SPX3P	(Chen et al. 2010)
			4	VfoOle		NA	PX5SPX3P	(Cao et al. 2014)
			5	VfoOle-GR218198	GR218198	NA	PX5SPX3P	(Chen et al. 2010)

*43 unique OLE, 33 OLE with perfect "Proline Knot" motif, 31 full-length OLEs.

NA, not available; NP, not published. Colored residues in "Proline Knot" column represents the deviation of the residue from the conserved "PX5SPX3P" motif. Highlighted names represent duplication or incomplete sequences.

amino acid residues in alignment have higher scores than those not identical and less similar residues. Each OLE sequence name is on the left of the alignment followed by the position of amino acid residue of OLE protein sequence in the alignment. The numbers at the top of the alignment are the positions of the multiple sequence alignment. The letters at the bottom of the alignment are the consensus residues. Color codes for amino acid residues are described in the figure legend.

Computational modeling

The secondary and three-dimensional structures were predicted with the I-TASSER method (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). Tung tree (*Vernicia fordii*) OLEs (VfoOLEs) were selected as examples of the five OLE subfamilies of tree OLEs for the modeling study. The secondary structure was predicted based on the formation of helix, strand, or coil, using a confident score of 0 to 9 assigned to individual amino acid residue. The higher score means the higher probability of the secondary structure. The solvent accessibility was predicted by assigning amino acid residue with a score of 0 (buried residue) to 9 (highly exposed residue). The normalized B-factor was predicted by assigning amino acid residue with a score of negative value (relatively more stable in the structure) or positive value (relatively unstable in the structure).

The top model of the three-dimensional structures of OLE was selected based on I-TASSER simulations which generate a large ensemble of structural conformations, called decoys. To select the top model, I-TASSER uses the SPICKER program to cluster all the decoys based on the pair-wise structure similarity, and reports up to five models which corresponds to the five largest structure clusters. The predicted ligand binding sites were generated by COACH based on the I-TASSER structure prediction. COACH is a meta-server approach that combines multiple function annotation results from the CO-FACTOR, TM-SITE, and S-SITE programs.

Results

Database search of OLEs

OLE sequences from tree species were obtained from keyword “oleosin” search and BlastP search against the National Center for Biotechnology Information (NCBI)’s non-redundant protein sequence databases (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). The complete genome sequences have been reported in five tree species. Forty-six OLE accessions were identified from the five trees whose genome sequences have been published in the literature and deposited in the GenBank database including 6 genes from *Prunus persica* (peach), 9 genes from *Populus trichocarpa* (poplar), 7 genes from *Ricinus communis* (castor bean), 8 genes from *Theobroma cacao* (cacao), and 16 genes from *Vitis vinifera* (grapevine) (Table 1).

In addition, 48 OLE accessions were identified from 17 other tree species in the GenBank database (Table 2). These OLEs included 1 OLE gene from *Coffea arabica* (coffee), 5 OLE genes from *Coffea canephora* (coffee), 2 OLE genes from *Corylus avellana* (hazelnut), 3 OLE genes from *Cocos nucifera* (coconut palm), 5 OLE genes from *Camellia oleifera* (tea oil), 5 OLE genes from *Cupressus sempervirens* (pencil pine), 1 OLE gene from *Citrus sinensis* (orange), 2 OLE gene from *Elaeis guineensis* (oil palm), 2 OLE genes

from *Ficus pumila* (climbing fig), 5 OLE genes from *Jatropha curcas* (barbados nut), 1 OLE gene from *Juglans regia* (walnut), 1 OLE gene from *Olea europaea* (olive), 1 OLE gene from *Persea americana* (avocado), 1 OLE gene from *Prunus armeniaca* (apricot), 5 OLE genes from *Prunus dulcis* (almond), 2 OLE genes from *Pinus taeda* (loblolly pine), and 5 OLE genes from *Vernicia fordii* (tung tree).

Unique OLE sequences

Duplication of OLE sequences from tree species in the GenBank database resulted from duplicated sequence projects and between the traditionally cloned genes and those derived from the completed genome sequences. Among the 46 OLEs accessions from the five tree genomes, two RcoOle and two TcaOle were duplicated, whereas seven VviOle were duplicated due to the two sequence projects. Elimination of the duplicated accessions reduced the total OLEs in the five sequenced genomes to 35 OLEs. In additions, two RcoOle consisted of PX5GPX3P and one TcaOle consisted of PX5SPX3C. Therefore, only thirty-two OLEs contained the perfect “Proline Knot” motif (PX5SPX3P) but one was incomplete (PtrOlep, no. 1) (Table 1). The full-length 31 OLEs were used for property and amino acid composition analysis later.

Among the 48 OLEs from 17 other species, only 43 OLEs were unique due to duplication of two JcuOle and five PduOle. Among them only 33 OLEs from 15 other tree species had the perfect “Proline Knot” motif (PX5SPX3P) because the sequences of three CnuOle and two CseOle were not completed with the first P missing, three CseOle contained SX5SIX3P and TX5SPX3T and one EguOle contained MX5SPX3P, and ParOle had only 29 amino acid residue. Finally, only 31 OLEs were full-length because 2 PtaOle were too short for property and amino acid composition analysis (Table 2). Taken together, 65 unique OLEs with the perfect “Proline Knot” motif (PX5SPX3P) were identified from 19 tree species, but only 62 full-length OLEs from 18 tree species were used for property and amino acid composition analysis.

Classification of OLE Isoforms

The complete genome sequences from the five trees provide a great opportunity for classifying tree OLEs. Phylogenetic analysis of these 32 OLEs with the perfect “Proline Knot” motif (PX5SPX3P) from the 5 sequenced genomes classified them into five well-clustered and one loose group of OLEs (Fig. 1A). All five trees were represented in the third, fourth, and fifth group of OLEs. However, no TcaOle was present in the first group of OLEs, and no VviOle was present in the second group of OLEs (Fig. 1A).

Instead, one TcaOle and three VviOle were found in the loose group at the bottom of the phylogenetic tree. Phylogenetic analysis of 33 unique OLEs from 14 other tree species resulted in a similar distribution of OLEs (Fig. 1B). As a comparison, seven maize OLEs were included in the phylogenetic analysis and showed a different pattern of clusters from those of the tree OLEs (Fig. 1B). These analyses suggest that there may be only five OLE subfamilies in tree species.

OLE properties and amino acid composition

It is generally observed that OLEs are small hydrophobic proteins. The properties and amino acid composition of 31

FIG. 1. Classification of 65 OLEs from 22 tree species and 7 OLEs from maize by phylogenetic analysis. The presumed evolutionary relationships among the OLEs were analyzed by phylogenetic analysis. (A) Phylogenetic analysis of 32 OLEs with the perfect “Proline Knot” motif (PX5SPX3P) from the five sequenced genomes. (B) Phylogenetic analysis of 33 unique OLEs from 14 other tree species and seven OLEs from maize. The numbers in the parenthesis following OLE names are the calculated distance values, which reflect the degree of divergence between all pairs of OLE sequences analyzed. The abbreviations of the organisms are: Car, *Coffea arabica* (coffee); Cca, *Coffea canephora* (coffee); Cav, *Corylus avellana* (hazelnut); Cnu, *Cocos nucifera* (coconut palm); Col, *Camellia oleifera* (tea oil); Cse, *Cupressus sempervirens* (pencil pine); Csi, *Citrus sinensis* (orange); Egu, *Elaeis guineensis* (oil palm); Fpu, *Ficus pumila* (climbing fig); Jcu, *Jatropha curcas* (barbados nut); Jre, *Juglans regia* (walnut); Oeu, *Olea europaea* (olive); Pam, *Persea Americana* (avocado); Par, *Prunus armeniaca* (apricot); Pdu, *Prunus dulcis* (almond); Ppe, *Prunus persica* (peach); Pta, *Pinus taeda* (loblolly pine); Ptr, *Populus trichocarpa* (poplar); Rco, *Ricinus communis* (castor bean); Tca, *Theobroma cacao* (cacao); Vfo, *Vernicia fordii* (tung tree); Vvi, *Vitis vinifera* (grapevine); Zma, *Zea mays* (maize). Figure 1 was modified from Figure 2A of a previous publication (Cao et al., 2014) with addition of 7 maize OLEs for better presentation of in-depth analysis of all tree oleosin sequences.

full-length OLEs from the five tree genomes were analyzed according to the classification of OLE subfamilies by phylogenetic analysis as described above. OLE1s had an average of 138 ± 6 amino acid residues ($n=7$) (Table 3). OLE2s had an average of 150 ± 8 amino acid residues ($n=7$) (Table 3). OLE3s had an average of 164 ± 36 amino acid residues ($n=5$) (Table 3). OLE4s had an average of 166 ± 57 amino acid residues ($n=6$) (Table 3). OLE5s had an average of 135 ± 15 amino acid residues ($n=6$) (Table 3).

OLE1s, OLE2s, OLE3s, OLE4s, and OLE5s had calculated molecular masses of 14.6, 15.8, 17.5, 17.9, and 14.4 kDa, respectively (Table 3). OLE1s, OLE2s, OLE3s, OLE4s, and OLE5s are all basic proteins with similar isoelectric points with an average value of 9.63, 9.12, 9.68, 9.09,

and 9.44, respectively (Table 3). They are all hydrophobic proteins with 42.97%, 42.39%, 39.32%, 42.73%, and 42.79% of the total residues being hydrophobic (AILEFWV), respectively (Table 3).

The properties and amino acid composition of all 62 OLEs after incorporation of the other 31 OLEs from 13 other trees were similar to those of the OLEs from the sequenced tree genomes (Table 4). OLE1s had an average of 140 ± 5 amino acid residues ($n=21$) (Table 4). OLE2s had an average of 155 ± 9 amino acid residues ($n=15$) (Table 4). OLE3s had an average of 158 ± 26 amino acid residues ($n=10$) (Table 4). OLE4s had an average of 186 ± 54 amino acid residues ($n=8$) (Table 4). OLE5s had an average of 139 ± 15 amino acid residues ($n=8$) (Table 4). OLE1s, OLE2s, OLE3s, OLE4s,

TABLE 3. PROPERTIES AND AMINO ACID COMPOSITION OF 31 OLEs FROM FIVE TREE GENOMES (% BY FREQUENCY)

OLE	Organism	Abbrev.	Ole name	Proline	Knot	Length (amino acid residue)	Molecular weight (Da)	Isoelectric point (PI)	Hydrophobic (AIFWV) (%)
1	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ19764.1	PX5SPX3P		148	15611.02	9.63	41.89
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002297927.1	PX5SPX3P		134	14083.89	9.87	44.03
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002304591.1	PX5SPX3P		134	14396.18	10.08	40.30
	<i>Ricinus communis</i> (castor bean)	Rco	RcoOle-XP_002511014.1	PX5SPX3P		138	14484.18	10.11	43.48
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY22772.1	PX5SPX3P		138	14449.84	8.37	44.93
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002281746.1	PX5SPX3P		144	15181.75	9.60	43.75
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_003631370.1	PX5SPX3P		133	13883.38	9.73	42.38
	Mean (n = 7)					138	14584.32	9.63	42.97
	SD (n = 7)					6	607.77	0.59	1.55
2	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ16302.1	PX5SPX3P		156	16625.60	9.98	39.74
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002300049.1	PX5SPX3P		149	15731.64	9.70	40.27
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002323804.1	PX5SPX3P		149	15855.77	10.00	40.94
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY02487.1	PX5SPX3P		160	16885.61	9.72	40.00
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002273158.2	PX5SPX3P		140	14753.63	9.16	44.29
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002275021.1	PX5SPX3P		157	16262.31	10.17	45.86
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002275087.1	PX5SPX3P		138	14330.89	5.11	45.65
	Mean (n = 7)					150	15777.92	9.12	42.39
	SD (n = 7)					8	942.64	1.80	2.76
3	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ10773.1	PX5SPX3P		223	23933.77	7.21	31.39
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002318659.1	PX5SPX3P		141	14392.19	11.00	41.84
	<i>Ricinus communis</i> (castor bean)	Rco	RcoOle-XP_002511342.1	PX5SPX3P		141	14955.00	10.96	42.55
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY20852.1	PX5SPX3P		175	19028.57	9.25	37.14
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002281381.1	PX5SPX3P		142	15007.05	9.98	43.66
	Mean (n = 5)					164	17463.32	9.68	39.32
	SD (n = 5)					36	4064.28	1.56	5.08
4	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ10882.1	PX5SPX3P		174	18770.12	9.93	38.51
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002317928.1	PX5SPX3P		164	17625.17	9.72	42.07
	<i>Ricinus communis</i> (castor bean)	Rco	RcoOle-XP_002511984.1	PX5SPX3P		169	18526.89	9.86	40.24
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY20922.1	PX5SPX3P		163	17529.00	9.61	43.56
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-CBI15408.3	PX5SPX3P		318	34680.14	10.02	35.85
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002273242.1	PX5SPX3P		161	17318.78	9.86	42.24
	Mean (n = 6)					166	17887.66	9.09	42.73
	SD (n = 6)					57	6293.26	1.35	3.80
5	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ25876.1	PX5SPX3P		125	13079.34	9.47	44.00
	<i>Populus trichocarpa</i> (poplar)	Ppe	PpeOle-EMJ26259.1	PX5SPX3P		111	11695.12	9.65	51.35
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002309527.1	PX5SPX3P		149	15729.24	9.59	38.26
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002334246.1	PX5SPX3P		134	14202.53	9.39	42.54
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY29822.1	PX5SPX3P		147	15803.16	9.34	39.46
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002275496.1	PX5SPX3P		146	15598.01	9.17	41.10
	Mean (n = 6)					135	14351.23	9.44	42.79
	SD (n = 6)					15	1688.56	0.17	4.67

Highlighted values were unusual properties of the OLE sequences.

TABLE 4. PROPERTIES AND AMINO ACID COMPOSITION OF 62 OLEs FROM 18 TREE SPECIES (% BY FREQUENCY)

OLE	Organism	Abbrev.	Ole name	Proline Knot	Length (amino acid residue)	Molecular weight (Da)	Isoelectric point (PI)	Hydrophobic (AIFWV) (%)
1	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ19764.1	PX5SPX3P	148	15611.02	9.63	41.89
	<i>Populus trichocarpa</i> (poplar)	Ppr	PprOle-XP_002297927.1	PX5SPX3P	134	14083.89	9.87	44.03
	<i>Populus trichocarpa</i> (poplar)	Ppr	PprOle-XP_002304591.1	PX5SPX3P	134	14396.18	10.08	40.30
	<i>Ricinus communis</i> (castor bean)	Rco	RcoOle-XP_002511014.1	PX5SPX3P	138	14484.18	10.11	43.48
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY22772.1	PX5SPX3P	138	14449.84	8.37	44.93
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002281746.1	PX5SPX3P	144	15181.75	9.60	43.75
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_003631370.1	PX5SPX3P	133	13883.38	9.73	42.38
	<i>Corylus avellana</i> (hazelnut)	Cav	CavOle-AAO65960.1	PX5SPX3P	140	14731.33	9.98	42.86
	<i>Juglans regia</i> (English walnut)	Jre	JreOle-AET74076.1	PX5SPX3P	140	14837.54	9.65	42.86
	<i>Juglans regia</i> (English walnut)	Jre	JreOle-AET74077.1	PX5SPX3P	139	14690.28	10.14	43.88
	<i>Camellia oleifera</i> (tea oil)	Col	ColOle-ABF57560.1	PX5SPX3P	141	14866.43	9.04	46.10
	<i>Coffea canephora</i> (coffee)	Cca	CcaOle-AAX49390.1	PX5SPX3P	135	14076.61	9.77	45.93
	<i>Prunus dulcis</i> (almond)	Pdu	PduOle-Q43804.1	PX5SPX3P	148	15611.02	9.63	41.89
	<i>Ficus pumila</i> (climbing fig)	Fpu	FpuOle-ABQ57397.1	PX5SPX3P	153	16314.13	9.77	42.48
	<i>Camellia oleifera</i> (tea oil)	Col	ColOle-ABF57563.1	PX5SPX3P	140	14813.18	9.52	42.14
	<i>Citrus sinensis</i> (sweet orange)	Csi	CsiOle-CAA88360.1	PX5SPX3P	144	15168.79	9.74	40.28
	<i>Jatropha curcas</i> (barbados nut)	Jcu	JcuOle-ABW90150.2	PX5SPX3P	137	14499.24	9.65	43.07
	<i>Jatropha curcas</i> (barbados nut)	Jcu	JcuOle-AFP19885.1	PX5SPX3P	137	14277.91	9.65	43.80
	<i>Vernicia fordii</i> (tung tree)	Vfo	VfoOle-GU245884	PX5SPX3P	137	14409.04	9.87	42.34
	<i>Coffea canephora</i> (coffee)	Cca	CcaOle-AAX49392.1	PX5SPX3P	142	15170.84	8.30	42.96
	<i>Elaeis guineensis</i> (oil palm)	Egu	EguOle-AF273023_1	PX5SPX3P	140	14634.21	8.28	45.00
	Mean (n = 21)				140	14770.99	9.54	43.16
	SD (n = 21)				5	584.70	0.56	1.54
2	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ16302.1	PX5SPX3P	156	16625.60	9.98	39.74
	<i>Populus trichocarpa</i> (poplar)	Ppr	PprOle-XP_002300049.1	PX5SPX3P	149	15731.64	9.70	40.27
	<i>Populus trichocarpa</i> (poplar)	Ppr	PprOle-XP_002323804.1	PX5SPX3P	149	15855.77	10.00	40.94
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY02487.1	PX5SPX3P	160	16885.61	9.72	40.00
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002273158.2	PX5SPX3P	140	14753.63	9.16	44.29
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002275021.1	PX5SPX3P	157	16262.31	10.17	45.86
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002275087.1	PX5SPX3P	138	14330.89	5.11	45.65
	<i>Corylus avellana</i> (hazelnut)	Cav	CavOle-AAO67349.2	PX5SPX3P	159	16697.64	10.54	40.88
	<i>Ficus pumila</i> (climbing fig)	Fpu	FpuOle-ABQ57396.1	PX5SPX3P	155	16535.16	9.85	39.35
	<i>Jatropha curcas</i> (barbados nut)	Jcu	JcuOle-ABW90149.2	PX5SPX3P	155	16607.91	9.99	41.94
	<i>Vernicia fordii</i> (tung tree)	Vfo	VfoOle-GU245885	PX5SPX3P	154	16787.83	9.93	38.96
	<i>Coffea canephora</i> (coffee)	Cca	CcaOle-AAX49391.1	PX5SPX3P	175	18672.79	9.07	38.29
	<i>Camellia oleifera</i> (tea oil)	Col	ColOle-ABF57561.1	PX5SPX3P	154	16143.25	9.69	40.91
	<i>Camellia oleifera</i> (tea oil)	Col	ColOle-ABF57562.1	PX5SPX3P	157	16688.52	9.37	40.13
	<i>Olea europaea</i> (olive)	Oeu	OeuOle-AAL92479.1	PX5SPX3P	165	17194.28	9.89	38.79
	Mean (n = 15)				155	16384.86	9.48	41.07
	SD (n = 15)				9	1010.06	1.27	2.39

(continued)

TABLE 4. (CONTINUED)

OLE	Organism	Abbrev.	Ole name	Proline Knot	Length (amino acid residue)	Molecular weight (Da)	Isoelectric point (PI)	Hydrophobic (AIFWV) (%)
3	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ10773.1	PX5SPX3P	223	23933.77	7.21	31.39
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002318659.1	PX5SPX3P	141	14392.19	11.00	41.84
	<i>Ricinus communis</i> (castor bean)	Rco	RcoOle-XP_002511342.1	PX5SPX3P	141	14955.00	10.96	42.55
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY20852.1	PX5SPX3P	175	19028.57	9.25	37.14
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002281381.1	PX5SPX3P	142	15007.05	9.98	43.66
	<i>Coffea arabica</i> (coffee)	Car	CarOle-AAY14574.1	PX5SPX3P	148	15684.69	9.87	47.30
	<i>Coffea canephora</i> (coffee)	Cca	CcaOle-AAX49389.1	PX5SPX3P	148	15744.75	9.87	45.95
	<i>Jatropha curcas</i> (barbados nut)	Jcu	JcuOle-ABW90148.2	PX5SPX3P	147	15580.55	10.28	40.14
	<i>Camellia oleifera</i> (tea oil)	Col	ColOle-ABF57559.1	PX5SPX3P	148	15902.76	9.84	41.22
	<i>Vernicia fordii</i> (tung tree)	Vfo	VfoOle-GR217754	PX5SPX3P	169	17963.62	8.25	34.32
	Mean (n = 10)				158	16819.30	9.65	40.55
	SD (n = 10)				26	2872.73	1.17	5.00
4	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ10882.1	PX5SPX3P	174	18770.12	9.93	38.51
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002317928.1	PX5SPX3P	164	17625.17	9.72	42.07
	<i>Ricinus communis</i> (castor bean)	Rco	RcoOle-XP_002511984.1	PX5SPX3P	169	18526.89	9.86	40.24
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY20922.1	PX5SPX3P	163	17529.00	9.61	43.56
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-CBI15408.3	PX5SPX3P	318	34680.14	10.02	35.85
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002273242.1	PX5SPX3P	161	17318.78	9.86	42.24
	<i>Coffea canephora</i> (coffee)	Cca	CcaOle-AAX49393.1	PX5SPX3P	167	17934.51	9.82	43.71
	<i>Vernicia fordii</i> (tung tree)	Vfo	VfoOle	PX5SPX3P	169	18288.73	9.17	44.38
	Mean (n = 8)				186	20084.17	9.75	41.32
	SD (n = 8)				54	5919.24	0.26	2.94
5	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ25876.1	PX5SPX3P	125	13079.34	9.47	44.00
	<i>Prunus persica</i> (peach)	Ppe	PpeOle-EMJ26259.1	PX5SPX3P	111	11695.12	9.65	51.35
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002309527.1	PX5SPX3P	149	15729.24	9.59	38.26
	<i>Populus trichocarpa</i> (poplar)	Ptr	PtrOle-XP_002334246.1	PX5SPX3P	134	14202.53	9.39	42.54
	<i>Theobroma cacao</i> (cacao)	Tca	TcaOle-EOY29822.1	PX5SPX3P	147	15803.16	9.34	39.46
	<i>Vitis vinifera</i> (grapevine)	Vvi	VviOle-XP_002275496.1	PX5SPX3P	146	15598.01	9.17	41.10
	<i>Persea americana</i> (avocado)	Pam	PamOle-AGT63296.1	PX5SPX3P	159	17469.57	9.74	42.14
	<i>Vernicia fordii</i> (tung tree)	Vfo	VfoOle-GR218198	PX5SPX3P	142	15105.86	9.52	42.25
	Mean (n = 8)				139	14835.35	9.48	42.64
	SD (n = 8)				15	1799.79	0.18	3.96

Highlighted values were unusual properties of the OLE sequences.

A

	(61)	61	70	80	90	100	110	120	130	142
PpeOle-EMJ10773	(46)	PSATHIVAMLT	TLVPIGGT	TLFSLGLT	LAGTTLGLAV	STPLFVIF	SPILVPAAL	VIGLSVVG	LTSCAFGIT	ALSSFSWLARF
PtrOle-XP_002318659	(21)	PSTSOVLAVV	TLISVGGT	TLFSLGLT	LVGTTLGLAV	TTPLFVIF	SPVLVPAAL	VIGLVLGL	FLASGAFG	VLTALSSLSWMA--
PtrOle-XP_002322186	(1)	-----	-----	-----	TLVGTTLGLAV	ATPLFVIF	SPALVPAAL	VIGLVLGL	FLASGAFG	VLTALSSLSWMA--
RcoOle-XP_002511342	(18)	PSTSOVLAVV	TLIPISCT	TLIFLAGAT	IVGALTGLG	VATPLFVIF	SPVLVPAAL	VIGLVLGL	FLASGAFG	VLTALSSLSWML--
PpeOle-EMJ16302	(27)	PTSGOVLAVIT	GLPVGGT	TLIALAGL	TLIMGTIIG	ALLATPLFI	IFSPVLVPAI	FVIGLAVT	GFITSGAF	GLTALSSLSWVT--
PtrOle-XP_002330049	(23)	PSASKALAVL	TMLPLGG	GLIALAG	TLIVGTTL	GLAVTTPL	FILESPVLV	PAALVIGL	AVTSFLAS	GAFGLTGSWSLSWVG--
PtrOle-XP_002323804	(23)	PSASKVLAVL	TMLPVGG	GLIALAG	TLIVGTTL	GLITVTTPL	FFLIFSPVLV	PAALLIG	FAATSFLAS	GAFGLTGFRLSWVA--
TcaOle-EOY02487	(31)	PSATQVLAVL	TLIPVGG	ILLALAG	TLITGTVIG	LVCATPLFI	IFSPVLVPAI	AVGLAVGL	FSSGAF	GLTGLSSLAYVE--
PpeOle-EMJ10882	(46)	NSTOLVGFL	TLIVSCA	ILLLTGL	TVATILGL	LIIVTPLII	ISPTIWPV	VGALLFLT	VAGFVSM	CGFGAAAVSWMYR--
TcaOle-EOY20922	(35)	NSAQLLGFL	TLFVSSA	SILTLTGL	TVTAALG	VIFMPLIV	ISSPIWI	PVGTVIF	VTAGFL	SA CGFGLVVMALSWMYR--
PtrOle-XP_002317928	(36)	NSSOLVGL	TLTISG	SILLLT	VITVTGS	VGLTTLPL	IIISPTIWP	IGIVLFF	VVAGFL	SFCGFGGLAVAGLSWVYK--
RcoOle-XP_002511384	(41)	NSTQVIGL	TLTISG	SILLLT	VITVTGS	VGLTTLPL	IIISPTIWP	IGIVLFF	VVAGFL	SFCGFGGLAVAGLSWVYK--
VviOle-CBI15408	(33)	NSTQFIGL	ALIVSG	CILIFLT	GLTITAT	VIGLIFFT	PLILISPTIWP	PAGTVL	FLLAAG	FLSMCGFGLAVLAGLSWIYK--
VviOle-XP_002273242	(33)	NSTQFIGL	ALIVSG	CILIFLT	GLTITAT	VIGLIFFT	PLILISPTIWP	PAGTVL	FLLAAG	FLSMCGFGLAVLAGLSWIYK--
PpeOle-EMJ19764	(23)	SYOVAKAAT	AVTAGGS	LLVLSGL	VLAGTVI	ALTATPL	LVIFSPVLV	PALITVAL	ITMGFL	TSGGFGVAAVTVLSWIYK--
VviOle-XP_002281746	(21)	PSYQAVKAAT	AATAGGS	LLVLSGL	TVAGTVI	ALTATPL	LVIFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
TcaOle-EOY22772	(20)	SHQVKAAT	AATAGGS	LLVLSGL	TVAGTVI	ALTATPL	LVIFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
PtrOle-XP_002297927	(16)	SSQIVKATTA	AVTAGGS	LLVLSGL	TVAGTVI	ALTATPL	LVIFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
PtrOle-XP_002304591	(16)	SHQVKAAT	AVTAGGS	LLVLSGL	TVAGTVI	ALTATPL	LVIFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
RcoOle-XP_002511014	(16)	VNOLVKAAT	AATAGGS	LLVLSGL	TVAGTVI	ALTATPL	LVIFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
VviOle-XP_003631370	(12)	LSYOLIKT	STAATIG	CSMVLSGL	TIAGTVI	ALVATPL	LVIFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
PpeOle-EMJ25876	(24)	SLRLTGKFL	TAGAAGT	TLVLSGL	TLTGTVM	ALIMATPL	VLFSPVLV	PAGIVM	LLAAGF	VFSGGCGAAMTVMLYLN--
PpeOle-EMJ26259	(11)	HTVTKLLT	AAIVCAT	TLFSLGL	TLTGTVM	ALIMATPL	VLFSPVLV	PAGIVM	LLAAGF	VFSGGCGAAMTVMLYLN--
PtrOle-XP_002309527	(20)	SRKAVK	FMTAGT	IGAAIVLSGL	TLTGTVM	ALVATPL	VLFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
PtrOle-XP_002334246	(13)	SRKAVIKF	MTAGT	IGAAIVLSGL	TLTGTVM	ALVATPL	VLFSPVLV	PAIAVFF	LVGLF	ASGGLAVAGLSWIYR--
TcaOle-EOY29822	(19)	SSRQAAKFL	TATLTCAT	TLFSLGL	TLTGTVM	ALIMATPL	VLFSPVLV	PAGIVM	LLAAGF	VFSGGCGAAMTVMLYLN--
VviOle-XP_002275496	(18)	SSRQAAKFL	TATLTCAT	TLFSLGL	TLTGTVM	ALIMATPL	VLFSPVLV	PAGIVM	LLAAGF	VFSGGCGAAMTVMLYLN--
TcaOle-EOY20852	(30)	PTSKVLAVV	TLIPVAG	TLIGLGL	SLVGTTL	GLAFALIF	FLIFSPVLV	PAALVIG	ASVAGFL	TSGAFGITGLSSLSWIVN--
VviOle-XP_002273158	(10)	PSASKILAV	IALPLG	CTLLLSGL	TLFVGS	MLGLAVATPL	FLIFSPVLV	PAALVIG	ASVAGFL	TSGAFGITGLSSLSWIVN--
VviOle-XP_002275021	(31)	PTSOVLAVI	ALPLAG	FLLLLAGL	TAFAGS	ALGLAVITPL	FLIFSPVLV	PATIAIGL	AVSVFT	SCAFGVTGLSSLSWIVN--
VviOle-XP_002275087	(22)	LSISQILAL	ITLPTG	CTLLIFGL	TFAAS	MLGLAVATPL	FLIFSPVLV	PAALVIG	ASVAGFL	TSGAFGITGLSSLSWIVN--
VviOle-XP_002281381	(17)	STSOVLAVI	TLFFVGG	ILVLSGL	TLTLAGT	LGLTVATPL	FLIFSPVLV	PAALVIG	ASVAGFL	TSGAFGITGLSSLSWIVN--
Consensus	(61)	SSSQVLA	LTLL	VGGTLLVLSGL	TLTGTVIGL	VATPLLVIF	SPVLVPAAI	VGLIV	G	FLASGGFGVTALSSLSWIY

B

		(52)	52	60	70	80	90	100	110	120	130	142	
CarOle-AAY14574	(24)	PSHSHVLA	VVILLPV	AGVLLGL	SGILLVGT	VIGLAVTT	PLFVIFSPILV	PAVFA	LGLALAG	---FLTSGAF	GITALSSLSWMLN--YIRLM		
CcaOle-AAX49389	(24)	PSHSHVLA	VVILLPV	AGVLLGL	SGILLVGT	VIGLAVTT	PLFVIFSPILV	PAVFA	LGLALAG	---FLTSGAF	GITALSSLSWMLN--YIRLM		
ColOle-ABF57559	(26)	PSKSCILAV	VILLPV	GGFLLLSGL	TLTGTTLGL	AVTTPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRRS			
JcuOle-ABW90148	(22)	GFSTSOILAV	VILLPV	SGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVIRG	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--NYIRRM		
VfoOle3-GR217899.1	(30)	GFSTSOILAV	VILLPV	SGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--NYIRRM		
CavOle-AAO67349	(30)	GPSAVKMA	MAVLAAL	PVGGTLL	ALAGLTL	ACSVIGL	LVTSPLFI	IFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--LRCA	
ColOle-ABF57561	(27)	PSPSKVLAV	MALLPL	GGTLLC	AGITL	TVGLTGL	AVTTPLFI	IFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--LRGV	
ColOle-ABF57562	(31)	PSPSKVLAV	MALLPL	GGTLLC	AGITL	TVGLTGL	AVTTPLFI	IFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--LRQA	
CcaOle-AAX49391	(34)	GPSAIVAV	VILLPV	GGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--FRQK		
OeuOle-AAL92479	(34)	GPSAIVAV	VILLPV	GGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--LQEV		
JcuOle-ABW90149	(28)	GPSAQKVLAV	VILLPV	GGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--LRQV		
VfoOle2-GU245885	(27)	RPSHSHVLA	VILLPV	GGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--VQEA		
FpuOle-ABQ57396	(28)	GPSAGKVLAV	VILLPV	GGTLLF	LAGITL	TGLIALAV	ATPLFVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--FKRT		
CcaOle-AAX49393	(37)	APNSTOLL	GFLTLIV	SGGILL	TLTGLT	TLTGLT	FLIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YRFRGH		
VfoOle4	(37)	APNSTOLL	GFLTLIV	SGGILL	TLTGLT	TLTGLT	FLIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YRFRGH		
PtaOle-AFG55850	(1)	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----		
PtaOle-AFG55851	(1)	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----		
PamOle-AGT63296	(14)	APNINCTA	IFLTA	AAVGGT	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
VfoOle5-GR218198	(16)	APPSHPT	VKFLTA	ATLGA	AFLLSGL	TLTGTV	IALMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
CcaOle-AAX49392	(11)	QHSSTQV	KTTTAV	AVGG	SILLLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
EguOle-APF273023_1	(10)	QTSQRVK	KGAAAT	IGSLLV	SGTLT	LAGTVIGL	AVTPLL	VIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRLM	
ColOle-ABF57560	(14)	EPSSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
CavOle-AAO65960	(14)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
JreOle-AET74076	(14)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
JreOle-AET74077	(14)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
JcuOle-ABW90150	(13)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
VfoOle1-GU245884	(13)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
CsiOle-CAA88360	(17)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
ColOle-ABF57563	(14)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
CcaOle-AAX49390	(8)	QPLSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
FpuOle-ABQ57397	(25)	HRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
PduOle-Q43804.1	(20)	QPRSHQV	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
ZmaOle-NP_001105338	(34)	QGMATTA	KAATAA	TAGGS	LLVLSGL	TLTAT	ATSLMAT	PLLVISP	ILVPT	---TFLVATG	---FTSSGGF	GAALSVMVAYMYRYATGKH	
ZmaOle-NP_001142571	(1)	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----		
ZmaOle-NP_001146985	(37)	GFTASQAL	TVAILFL	PGGLLL	VLSGL	ALASTV	GLATPL	VELLSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRLM	
ZmaOle-NP_001146998	(47)	PPAPMA	VAKAAT	ATAAF	SMLLSGL	AVTGL	VALVATPL	LVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRLM	
ZmaOle-NP_001147032	(42)	GFTASQAL	TVAILFL	PGGLLL	VLSGL	ALASTV	GLATPL	VELLSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRLM	
ZmaOle-NP_001152686	(33)	SPRST	VATAAL	VPLGA	ALLRAS	GTALAA	TVTGLAT	PLVIFSPVLV	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRLM	
ZmaOle-NP_001152711	(27)	APNSTQV	VGFLTL	VSGAV	LLLTG	VLTGAV	VALVGL	GLALIT	SPVIRG	PAALVIG	LAVAG	---FLTSGAF	GITALSSLSWMLN--YIRLM
Consensus	(52)	P	S	QVL	LT	L	VGGSLLVLSGL	TLTGTVIGL	VATPLLVIF	SPVLVPAAI	VGLIV	G	

and OLE5s had calculated molecular masses of 14.8, 16.4, 16.8, 20.1, and 14.8 kDa, respectively (Table 4). OLE1s, OLE2s, OLE3s, OLE4s and OLE5s are all basic proteins with similar isoelectric points with an average value of 9.54, 9.48, 9.65, 9.75, and 9.48, respectively (Table 4). They are all hydrophobic proteins with 43.16%, 41.07%, 40.55%, 41.32%, and 42.64% of the total residues being hydrophobic (AILFWV), respectively (Table 4).

Conservation of the “proline knot” among all OLEs

Multiple sequence alignment was performed initially to analyze all 32 unique OLEs from the five sequenced genomes of tree species. Sequence alignment identified only four completely conserved amino acid residues among the 32 OLEs. All of them were in the “proline knot” motif (Fig. 2A). These four residues in the “proline knot” motif were also completely conserved in 33 unique OLEs from the other 14 tree species and all seven maize OLEs (Fig. 2B). However, OLEs of some Genbank accessions possessed divergent residues in the “proline knot” motif. Two RcoOle consisted of PX5GPX3P and one TcaOle consisted of PX5SPX3C (Table 1). Three CseOle contained SX5SIX3P and TX5SPX3T and one EguOle contained MX5SPX3P (Table 2). It is unknown if they reflect true deviations or sequence errors.

Identification of amino acid residues and sequence motifs conserved in OLE subfamilies

Multiple sequence alignment was performed to identify conserved amino acid residues and sequence motifs within OLE subfamily according to the classification of OLEs by phylogenetic analysis using OLE sequences from the sequenced genomes. Besides the four residues completely conserved in all but one with PX5GPX3P (RcoOle) of the 32 OLEs, numerous residues are conserved within the subfamilies (Fig. 3). Many unique motifs could be identified among the OLE subfamilies based on the sequenced genomes: 45 residues in OLE1s ($n=7$), 35 residues in OLE2s ($n=7$), 40 in OLE3s ($n=5$), 70 in OLE4s ($n=6$), 39 in OLE5s ($n=6$) are conserved (Fig. 3). The number of conserved residues were reduced when all 62 OLEs from 18 tree species (the two PtaOle and one PtrOle were not included because of the incomplete sequences) were used in the alignment: 27 out of 140 residues in OLE1s (19%) ($n=21$), 19 out of 155 residues in OLE2s (12%) ($n=15$), 29 out of 158 residues in OLE3s (18%) ($n=10$), 58 out of 186 residues in OLE4s (31%) ($n=8$) and 31 out of 139 residues in OLE5s (22%) ($n=8$) (Fig. 4).

There were distinct patterns of conservation among the five OLE subfamilies. The 27 completely conserved residues among OLE1s were distributed evenly with 10 residues before and 11 residues after the “proline knot” motif (6 residues); while eight of the conserved residues were present at the N-terminal regions of OLE2s and only 4 residues after “proline knot” motif of OLE2 subfamilies (7 residues) (Fig. 4). Among the OLE3s, the 29 completely conserved residues were not distributed evenly with 14 residues before and 10 residues including one 5-consecutive residue motif after the “proline knot” motif (5 residues) (Fig. 4). OLE4s were the most conserved OLE subfamilies with 15 of the conserved residues present at the N-terminal regions before “proline knot” motif (9 residues) and 34 residues including a 22-residue highly conserved motif out of 30 residues at the C-termini of the proteins (Fig. 4). In contrast to OLE4 subfamilies, the N-termini of OLE5s were more conserved than the C-termini of the proteins with 17 residues including a 7-residues consecutively and 8 conserved residues before and after the “proline knot” motif (6 residues) (Fig. 4).

Identification of amino acid residues and sequence motifs conserved within tree species

It is not clear why different plant species accumulate different TAGs in terms of the amount and types of TAGs. A comparison of OLEs based on species might point out some useful features. Among the 19 tree species, 9 trees have almost all OLEs identified. These OLEs provide an opportunity of identifying conserved residues within a species. Multiple sequence alignment showed many conserved residues are species-specific (Fig. 5).

Peach OLEs have at least 19 residues completely conserved in the 6 OLEs with 10 residue before and 4 residue after the “proline knot” motif (5 residues). Poplar OLEs have at least 18 conserved residues with 8 before and 6 after the “proline knot” motif (4 residues). Castor bean OLEs have 26 conserved residues with 12 before and 9 after the “proline knot” motif (5 residues). Cacao tree OLEs have 19 conserved residues with 8 before and 6 after the “proline knot” motif (5 residues). Grape OLEs have at least 13 conserved residues with 6 before and 2 after the “proline knot” motif (5 residues). Coffee tree OLEs have at least 26 conserved residues with 11 before and 8 after the “proline knot” motif (7 residues). Tea oil tree OLEs have at least 31 conserved residues with 11 before and 12 after the “proline knot” motif (8 residues). Barbados nut tree OLEs have 36 conserved residues with 16

FIG. 2. Identification of amino acid residues conserved in all OLEs by multiple sequence alignment. Multiple sequence alignment was performed using the ClustalW algorithm of the AlignX program of the Vector NTI software. (A) Multiple sequence alignment of 32 OLEs with the perfect “Proline Knot” motif (PX5SPX3P) from the five sequenced genomes. (B) Multiple sequence alignment of 33 unique OLEs from 14 other tree species and seven OLEs from maize. Each OLE sequence name is on the left of the alignment, followed by the position of amino acid residue of OLE protein sequence in the alignment. The letters at the bottom of the alignment are the consensus residues. Color codes for amino acid residues are as follows: 1) red on yellow: consensus residue derived from a completely conserved residue at a given position; 2) blue on cyan: consensus residue derived from the occurrence of greater than 50% of a single residue at a given position; 3) black on white: nonsimilar residues. The abbreviations of the organisms are described in Figure 1 legend. Figure 2 was modified from Supplementary Figure S2 of a previous publication (Cao et al., 2014), with addition of 7 maize OLEs for better presentation of in-depth analysis of all tree oleosin sequences.

TreeGenomeOlePSP-31 OLEs from 5 tree species-Classification-5 subfamilies

FIG. 3. Identification of amino acid residues and sequence motifs conserved in OLEs from five sequenced tree genomes. Multiple sequence alignment was performed using the ClustalW algorithm and 31 OLE protein sequences from 5 trees (listed in Table 1). The color code and related information are described in “Methods” section.

TreeOlePSP-62 OLEs from 18 tree species-Classification-5 subfamilies

FIG. 4. Identification of amino acid residues and sequence motifs conserved in 62 OLEs from 18 tree species.

[illegible]

FIG. 5. Identification of amino acid residues and sequence motifs conserved in OLEs within each tree.

(Continued next page)

C Predicted normalized B-factor

VfoOLE1

VfoOLE2

VfoOLE3

VfoOLE4

VfoOLE5

FIG. 6. (Continued).

VfoOLE1

VfoOLE2

FIG. 7. (Continued next page).

before and 12 after the “proline knot” motif (8 residues). Tung oil tree OLEs have 16 conserved residues with 8 before and 4 after the “proline knot” motif (4 residues).

Prediction of the secondary structures of the five OLEs from tung tree

To gain structural information for tree OLEs, tung tree (*Vernicia fordii*) OLEs (VfoOLEs) representing the five OLE subfamilies were selected for structural analyses using computational modeling. The I-TASSER program predicted that VfoOLE1 consisted of three long helixes and a short helix connected with short coil sequences, but without any strand in the secondary structure (Fig. 6A). Similarly, the other four VfoOLEs contained at least three helixes connected with short coils and no strand. VfoOLE2, VfoOLE3, VfoOLE4 and VfoOLE5 contained three, five, four, and four helixes, respectively (Fig. 6A). These helix spanning structures were predicted to be buried in the lipid membranes (Fig. 6B), and contributed to the stability of the proteins (Fig. 6C).

Prediction of the three-dimensional structures of the five OLEs from tung tree

Computational modeling was also used to predict the three-dimensional structures of tung tree OLEs. The top models showed distinct structures among VfoOLEs and the predict ligand binding sites were also different among themselves (Fig. 7). VfoOLE1 contained numerous ligand binding sites, compared the other four VfoOLEs with only a few potential ligand binding sites (Fig. 7). VfoOLE3 displayed a very different structure from the other VfoOLEs because the five helixes in VfoOLE3 were arranged in two long parallel strands (Fig. 7).

Discussion

Triacylglycerols (TAGs) are one of the major energy storage molecules in eukaryotic organisms. They provide fatty acids for membrane biogenesis and lead to obesity if excessive amounts of fats accumulate in adipose tissues (Chapman et al., 2012; Dyer et al., 2008; Farese, Jr. and

VfoOLE3

VfoOLE4

FIG. 7. (Continued next page).

Walther, 2009; Somerville et al., 2000). Plant TAGs are both for human consumption and for industrial applications (Cahoon et al., 2007; Dyer et al., 2008) including biofuel production (Carvalho et al., 2008; Chen et al., 2010b; Misra et al., 2013; Tai and Stephanopoulos, 2013; Winichayakul et al., 2013). Many trees contain up to 50%–80% of oil in the seeds and fruits (Cao et al., 2013; Chen et al., 2008; Zeng et al., 2014; Zhang et al., 2014). The amounts of TAGs are influenced by OLEs, generally low-molecular-mass hydrophobic proteins covering the oil bodies/droplets (Dyer et al., 2008; Huang, 1992). However, the precise functions of OLEs are to be determined, and there is a lack of comprehensive sequence analysis and structural information of OLEs among diverse trees.

This study identified OLEs from tree species, performed genome-wide analysis of tree OLEs, classified tree OLEs, identified conserved sequence motifs and amino acid residues, and predicted secondary and three-dimensional structures in tree OLEs and OLE subfamilies. Data mining identified 65 OLEs with perfect conservation of the “proline knot” motif (PX5SPX3P) from 19 tree species. These OLEs

had similar properties and amino acid composition with >40% hydrophobic amino acid residues.

Genome-wide phylogenetic analysis and multiple sequence alignment separated these proteins into five OLE subfamilies. There were distinct patterns of sequence conservation among the OLE subfamilies and within individual species. Computational modeling predicted OLEs contained at least three helices connected with short coils but without any strand. These helices may provide stable structures and tightly bind to TAGs on the surface of oil bodies. The three-dimensional modeling displayed distinct structure patterns with different ligand binding sites among the five VfoOLEs.

The systematic sequence analyses of OLE sequences led to identify incomplete OLE sequences, repetitive entries in the database, and divergent sequences from the completely conserved motif. These analyses provide fundamental sequence and structural information for diverse OLE isoforms within the same subfamily and among the different species, which should facilitate studying the structure-function relationship and identify critical amino acid residues in OLEs.

VfoOLE5

FIG. 7. Predicted three-dimensional structures and the potential ligand binding sites of the five tung tree OLEs. The *top model* of the three-dimensional structures of VfoOLE was selected based on I-TASSER simulations. The predicted ligand binding sites were generated by COACH based on the I-TASSER structure prediction.

This agrigenomic information should be the basis for engineering OLEs designated to increase oil content and diversify oil varieties to meet the food, feed, and biofuels challenges in the 21st century.

Acknowledgments

This work was supported by USDA-ARS Quality and Utilization of Agricultural Products Research Program 306 through CRIS 6054-41000-103-00D.

Data Archiving Statement: All sequence data mentioned in this manuscript are extracted from the National Center for Biotechnology Information (NCBI)'s non-redundant protein sequence databases (<http://blast.ncbi.nlm.nih.gov/blast.cgi>), which are accessible to the public.

Author Disclosure Statement

The authors declare that no conflicting financial interests exist.

References

- Akkerdaas JH, Schocker F, Vieths S, et al. (2006). Cloning of oleosin, a putative new hazelnut allergen, using a hazelnut cDNA library. *Mol Nutr Food Res* 50, 18–23.
- Altschul SF, Gish W, Miller W, Myers EW, and Lipman DJ. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403–410.
- Altschul SF, and Lipman DJ. (1990). Protein database searches for multiple alignments. *Proc Natl Acad Sci USA* 87, 5509–5513.
- Beatty MK, Rahman A, Cao H, Woodman W, Lee M, Myers AM, and James MG. (1999). Purification and molecular genetic characterization of ZPU1, a pullulanase-type starch-debranching enzyme from maize. *Plant Physiol* 119, 255–266.
- Cahoon EB, Shockey JM, Dietrich CR, Gidda SK, Mullen RT, and Dyer JM. (2007). Engineering oilseeds for sustainable production of industrial and nutritional feedstocks: solving bottlenecks in fatty acid flux. *Curr Opin Plant Biol* 10, 236–244.
- Cao H. (2011). Structure-function analysis of diacylglycerol acyltransferase sequences from 70 organisms. *BMC Res Notes* 4, 249.
- Cao H, Imparl-Radosevich J, Guan H, Keeling PL, James MG, and Myers AM. (1999). Identification of the soluble starch synthase activities of maize endosperm. *Plant Physiol* 120, 205–216.
- Cao H, Shockey JM, Klasson KT, Mason CB, and Scheffler BE. (2013). Developmental regulation of diacylglycerol acyltransferase family gene expression in tung tree tissues. *PLoS ONE* 8, e76946.
- Cao H, Zhang L, Tan X, Long H, and Shockey JM. (2014). Identification, classification and differential expression of oleosin genes in tung tree (*Vernicia fordii*). *PLoS ONE* 9, e88409.
- Carvalho CR, Clarindo WR, Praca MM, Araujo FS, and Carels N. (2008). Genome size, base composition and karyotype of *Jatropha curcas* L., an important biofuel plant. *Plant Sci* 174, 613–617.
- Chan AP, Crabtree J, Zhao Q, et al. (2010). Draft genome sequence of the oilseed species *Ricinus communis*. *Nat Biotechnol* 28, 951–956.
- Chapman KD, Dyer JM, and Mullen RT. (2012). Biogenesis and functions of lipid droplets in plants: Thematic Review Series: Lipid Droplet Synthesis and Metabolism: from Yeast to Man. *J Lipid Res* 53, 215–226.
- Chen Y, Wang Y, Zhou G, Li P, and Zhang S. (2008). Key mediators modulating TAG synthesis and accumulation in woody oil plants. *Afr J Biotechnol* 7, 4743–4749.
- Chen Y, Zhou G, Wang Y, and Xu L. (2010a). F-BOX and oleosin: Additional target genes for future metabolic engineering in tung trees? *Ind Crops Prod* 32, 684–686.
- Chen YH, Chen JH, Chang CY, and Chang CC. (2010b). Biodiesel production from tung (*Vernicia montana*) oil and its blending properties in different fatty acid compositions. *Bioresour Technol* 101, 9521–9526.
- Chua AC, Jiang PL, Shi LS, Chou WM, and Tzen JT. (2008). Characterization of oil bodies in jelly fig achenes. *Plant Physiol Biochem* 46, 525–532.

- Cummins I, and Murphy DJ. (1992). cDNA sequence of a sunflower oleosin and transcript tissue specificity. *Plant Mol Biol* 19, 873–876.
- Dyer JM, Stymne S, Green AG, and Carlsson AS. (2008). High-value oils from plants. *Plant J* 54, 640–655.
- Eastmond PJ. (2004). Cloning and characterization of the acid lipase from castor beans. *J Biol Chem* 279, 45540–45545.
- Farese RV Jr, and Walther TC. (2009). Lipid droplets finally get a little R-E-S-P-E-C-T. *Cell* 139, 855–860.
- Garcia-Mas J, Messegueur R, Arus P, and Puigdomenech P. (1995). Molecular characterization of cDNAs corresponding to genes expressed during almond (*Prunus amygdalus* Batsch) seed development. *Plant Mol Biol* 27, 205–210.
- Geuna F, Banfi R, and Bassi D. (2005). Identification and characterization of transcripts differentially expressed during development of apricot (*Prunus armeniaca* L.) fruit. *Tree Genet Genomes* 1, 69–78.
- Giannoulia K, Banilas G, and Hatzopoulos P. (2007). Oleosin gene expression in olive. *J Plant Physiol* 164, 104–107.
- Gu K, Yi C, Tian D, Sangha J, Hong Y, and Yin Z. (2012). Expression of fatty acid and lipid biosynthetic genes in developing endosperm of *Jatropha curcas*. *Biotechnol Biofuels* 5, 47.
- Guilloteau M, Laloi M, Blais D, Crouzillat D, and McCarthy J. (2003). Oil bodies in *Theobroma cacao* seeds: Cloning and characterization of cDNA encoding the 15.8 and 16.9 kDa oleosins. *Plant Sci* 164, 597–606.
- Hernandez-Pinzon I, Ross JH, Barnes KA, Damant AP, and Murphy DJ. (1999). Composition and role of tapetal lipid bodies in the biogenesis of the pollen coat of *Brassica napus*. *Planta* 208, 588–598.
- Hu Z, Wang X, Zhan G, Liu G, Hua W, and Wang H. (2009). Unusually large oilbodies are highly correlated with lower oil content in *Brassica napus*. *Plant Cell Rep* 28, 541–549.
- Huang AH. (1996). Oleosins and oil bodies in seeds and other organs. *Plant Physiol* 110, 1055–1061.
- Huang AHC. (1992). Oil bodies and oleosins in seeds. *Annu Rev Plant Physiol Plant Mol Biol* 43, 177–200.
- Hyun TK, Kumar D, Cho YY, Hyun HN, and Kim JS. (2013). Computational identification and phylogenetic analysis of the oil-body structural proteins, oleosin and caleosin, in castor bean and flax. *Gene* 515, 454–460.
- Jaillon O, Aury JM, Noel B, et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467.
- Jolivet P, Acevedo F, Boulard C, et al. (2013). Crop seed oil bodies: From challenges in protein identification to an emerging picture of the oil body proteome. *Proteomics* 13, 1836–1849.
- Kalinski A, Loer DS, Weisemann JM, Matthews BF, and Herman EM. (1991). Isoforms of soybean seed oil body membrane protein 24 kDa oleosin are encoded by closely related cDNAs. *Plant Mol Biol* 17, 1095–1098.
- Keddie JS, Edwards EW, Gibbons T, Shaw CH, and Murphy DJ. (1992a). Sequence of an oleosin cDNA from *Brassica napus*. *Plant Mol Biol* 19, 1079–1083.
- Keddie JS, Hubner G, Slocombe SP, et al. (1992b). Cloning and characterisation of an oleosin gene from *Brassica napus*. *Plant Mol Biol* 19, 443–453.
- Kovach A, Wegrzyn JL, Parra G, et al. (2010). The *Pinus taeda* genome is characterized by diverse and highly diverged repetitive sequences. *BMC Genomics* 11, 420.
- Larkin MA, Blackshields G, Brown NP, et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.
- Lin X, Kaul S, Rounsley S, et al. (1999). Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402, 761–768.
- Long H, Tan X, Chen H, Zhang L, and Hu J. (2010). Cloning and sequence analysis of full-length cDNAs encoding oleosins from *Vernicia fordii*. *J Cent South Univ Forestry Technol* 30, 31–38.
- Lopez-Gomez R, Torres-Cardenas Y, Chavez-Moctezuma M, et al. Genomics of avocado criollo fruit (*Persea americana* Mill. VAR. DRYMIFOLIA). *Actas VI Congreso Mundial del Aguacate. Proceedings VI World Avocado Congress (Actas VI Congreso Mundial del Aguacate) 1-10, 2007. 11-12-2007.*
- Mayer K, Schuller C, Wambutt R, et al. (1999) Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*. *Nature* 402, 769–77.
- Mayfield JA, and Preuss D. (2000). Rapid initiation of *Arabidopsis* pollination requires the oleosin-domain protein GRP17. *Nat Cell Biol* 2, 128–130.
- Misra N, and Panda PK. (2013). In search of actionable targets for agrigenomics and microalgal biofuel production: Sequence-structural diversity studies on algal and higher plants with a focus on GPAT protein. *OMICS* 17, 173–186.
- Misra N, Panda PK, and Parida BK. (2013). Agrigenomics for microalgal biofuel production: An overview of various bioinformatics resources and recent studies to link OMICS to bioenergy and bioeconomy. *OMICS* 17, 537–549.
- Motamayor J, Mockaitis K, Schmutz J, et al. (2013). The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol* 14, R53.
- Naot D, Holland D, Avsian-Kretchmer O, Eshdat Y, and Ben-Hayyim G. (1995). Induction of a gene encoding an oleosin homologue in cultured citrus cells exposed to salt stress. *Gene* 161, 171–173.
- Parthibane V, Iyappan R, Vijayakumar A, Venkateshwari V, and Rajasekharan R. (2012a). Serine/threonine/tyrosine protein kinase phosphorylates oleosin, a regulator of lipid metabolic functions. *Plant Physiol* 159, 95–104.
- Parthibane V, Rajakumari S, Venkateshwari V, Iyappan R, and Rajasekharan R. (2012b). Oleosin is bifunctional enzyme that has both monoacylglycerol acyltransferase and phospholipase activities. *J Biol Chem* 287, 1946–1954.
- Pastor S, Sethumadhavan K, Ullah AHJ, et al. (2013). Molecular properties of the class III subfamily of acyl-coenzyme A binding proteins from tung tree (*Vernicia fordii*). *Plant Science* 203–204, 79–88.
- Pedron L, Baldi P, Hietala AM, and La PN. (2009). Genotype-specific regulation of cold-responsive genes in cypress (*Cupressus sempervirens* L.). *Gene* 437, 45–53.
- Popluechai S, Froissard M, Jolivet P, Breviaro D, Gatehouse AM, O'Donnell AG, Chardot T, and Kohli A. (2011). *Jatropha curcas* oil body proteome and oleosins: L-form JcOle3 as a potential phylogenetic marker. *Plant Physiol Biochem* 49, 352–356.
- Qu RD, and Huang AH. (1990). Oleosin KD 18 on the surface of oil bodies in maize. Genomic and cDNA sequences and the deduced protein structure. *J Biol Chem* 265, 2238–2243.
- Regalado ES, Angeles JGC, Rodelas AJD, Garcia RN, Laurena AC, and Mendoza EMT. (2008). Cloning and characterization of the oleosin cDNA isoforms from Coconut (*Cocos nucifera* L.). *Philippine Agricult Sci* 91, 37.
- Rooijen G, Terning L, and Moloney M. (1992). Nucleotide sequence of an *Arabidopsis thaliana* oleosin gene. *Plant Mol Biol* 18, 1177–1179.

- Roy A, Kucukural A, and Zhang Y. (2010). I-TASSER: A unified platform for automated protein structure and function prediction. *Nat Protoc* 5, 725–738.
- Sahu J, Sen P, Choudhury MD, Dehury B, Barooah M, Modi MK, and Talukdar AD. (2014). Rediscovering medicinal plants' potential with OMICS: Microsatellite survey in expressed sequence tags of eleven traditional plants with potent antidiabetic properties. *OMICS* 18, 298–309.
- Saitou N, and Nei M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406–425.
- Salanoubat M, Lemcke K, Rieger M, et al. (2000). Sequence and analysis of chromosome 3 of the plant *Arabidopsis thaliana*. *Nature* 408, 820–822.
- Sarmiento C, Ross JH, Herman E, and Murphy DJ. (1997). Expression and subcellular targeting of a soybean oleosin in transgenic rapeseed. Implications for the mechanism of oil-body formation in seeds. *Plant J* 11, 783–796.
- Schmidt MA, and Herman EM. (2008). Suppression of soybean oleosin produces micro-oil bodies that aggregate into oil body/ER complexes. *Mol Plant* 1, 910–924.
- Shimada TL, Shimada T, Takahashi H, Fukao Y, and Hara-Nishimura I. (2008). A novel role for oleosins in freezing tolerance of oilseeds in *Arabidopsis thaliana*. *Plant J* 55, 798–809.
- Siloto RM, Findlay K, Lopez-Villalobos A, Yeung EC, Nykiforuk CL, and Moloney MM. (2006). The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. *Plant Cell* 18, 1961–1974.
- Simkin AJ, Qian T, Caillet V, et al. (2006). Oleosin gene family of *Coffea canephora*: quantitative expression analysis of five oleosin genes in developing and germinating coffee grain. *J Plant Physiol* 163, 691–708.
- Somerville C, Browse J, Jaworski JG, and Ohlrogge JB. (2000). Lipids. In: B Buchanan, W Gruissem, R Jones, eds, *Biochemistry and Molecular Biology of Plants*. American Society of Plant Physiologists, Rockville, MD, pp. 456–527.
- Tabata S, Kaneko T, Nakamura Y, et al. (2000). Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. *Nature* 408, 823–826.
- Tai M, and Stephanopoulos G. (2013). Engineering the push and pull of lipid biosynthesis in oleaginous yeast *Yarrowia lipolytica* for biofuel production. *Metab Eng* 15, 1–9.
- Theologis A, Ecker JR, Palm CJ, et al. (2000). Sequence and analysis of chromosome 1 of the plant *Arabidopsis thaliana*. *Nature* 408, 816–820.
- Thompson JD, Higgins DG, and Gibson TJ. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22, 4673–4680.
- Ting JT, Lee K, Ratnayake C, Platt KA, Balsamo RA, and Huang AH. (1996). Oleosin genes in maize kernels having diverse oil contents are constitutively expressed independent of oil contents. Size and shape of intracellular oil bodies are determined by the oleosins/oils ratio. *Planta* 199, 158–165.
- Tuskan GA, Difazio S, Jansson S, et al. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* 313, 1596–1604.
- Vance VB, Huang AH. (1987). The major protein from lipid bodies of maize. Characterization and structure based on cDNA cloning. *J Biol Chem* 262, 11275–11279.
- Verde I, Abbott AG, Scalabrin S, et al. (2013). The high-quality draft genome of peach (*Prunus persica*) identifies unique patterns of genetic diversity, domestication and genome evolution. *Nat Genet* 45, 487–494.
- Walther TC, and Farese RV, Jr. (2012). Lipid droplets and cellular lipid metabolism. *Annu Rev Biochem* 81, 687–714.
- Wang X, McLachlan J, Zamore PD, and Hall TM. (2002). Modular recognition of RNA by a human pumilio-homology domain. *Cell* 110, 501–512.
- Wang X, and Tanaka Hall TM. (2001). Structural basis for recognition of AU-rich element RNA by the HuD protein. *Nat Struct Biol* 8, 141–145.
- Wang X, Zamore PD, and Hall TM. (2001). Crystal structure of a Pumilio homology domain. *Mol Cell* 7, 855–865.
- Winichayakul S, Scott RW, Roldan M, et al. (2013). In vivo packaging of triacylglycerols enhances *Arabidopsis* leaf biomass and energy density. *Plant Physiol* 162, 626–639.
- Xu R, Wang R, and Liu A. (2011). Expression profiles of genes involved in fatty acid and triacylglycerol synthesis in developing seeds of *Jatropha* (*Jatropha curcas* L.). *Biomass Bioenerg* 35, 1683–1692.
- Yang J, Yan R, Roy A, Xu D, Poisson J, and Zhang Y. (2015). The I-TASSER Suite: Protein structure and function prediction. *Nat Methods* 12, 7–8.
- Yong S, Yusuf R, Koh CL, Cheah SC, and Ooi LCL. (1999). Partial nucleotide sequence of an oleosin cDNA from *Elaeis guineensis* Jacq. *J Oil Palm Res* 11, 18–23.
- Zeng Y, Tan X, Zhang L, Jiang N, and Cao H. (2014). Identification and expression of fructose-1,6-bisphosphate aldolase genes and their relations to oil content in developing seeds of tea oil tree (*Camellia oleifera*). *PLoS ONE* 9, e107422.
- Zhang L, Jia B, Tan X, et al. (2014). Fatty acid profile and unigene-derived simple sequence repeat markers in tung tree (*Vernicia fordii*). *PLoS ONE* 9, e105298.
- Zhang Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinform* 9, 40.
- Zimmerman JM, Eliezer N, and Simha R. (1968). The characterization of amino acid sequences in proteins by statistical methods. *J Theor Biol* 21, 170–201.

Address correspondence to:

Dr. Heping Cao
U.S. Department of Agriculture
Agricultural Research Service
Southern Regional Research Center
1100 Robert E. Lee Blvd
New Orleans, Louisiana 70124
E-mail: Heping.Cao@ars.usda.gov