

Molecular Characterization and Haplotype Analysis of *Low Phytic Acid-1 (lpa1)* Gene Governing Accumulation of Kernel Phytic Acid in Subtropically-Adapted Maize

Vinay Bhatt ^{1,2}, Vignesh Muthusamy ^{1,*}, Rashmi Chhabra ¹, Ashvinkumar Katral ¹, Shridhar Ragi ¹, Vinay Rojaria ¹, Gulab Chand ¹, Govinda Rai Sarma ¹, Rajkumar Uttamrao Zunjare ¹, Kusuma Kumari Panda ², Ashok Kumar Singh ¹ and Firoz Hossain ¹

¹ Division of Genetics, ICAR-Indian Agricultural Research Institute, New Delhi 110012, India; vinaybhatt024@gmail.com (V.B.); resh0428@rediffmail.com (R.C.); ashokgkatral@gmail.com (A.K.); shridharragi1996@gmail.com (S.R.); vvinyay3878@yahoo.com (V.R.); gulab.biotech@yahoo.com (G.C.); raisarmag@gmail.com (G.R.S.); raj_gpb@yahoo.com (R.U.Z.); aks_gene@yahoo.com (A.K.S.); fh_gpb@yahoo.com (F.H.)

² AMITY Institute of Biotechnology, AMITY University, Noida, Uttar Pradesh 201313, India; pkkumari@amity.edu

* Correspondence: pmvignesh@yahoo.co.in

SUPPLEMENTARY MATERIAL

Table S1. Information on the 48 maize inbreds utilised in an *InDel*-based gene diversity analysis (WT: wild-type; MT: mutant-type)

S. No.	Inbreds	Type	Source
1	UMI1200	WT	TNAU, Coimbatore
2	UMI1230	WT	TNAU, Coimbatore
3	V335	WT	ICAR-VPKAS, Almora
4	V373	WT	ICAR-VPKAS, Almora
5	BML6	WT	PJTSAU, Telangana
6	BML7	WT	PJTSAU, Telangana
7	BML45	WT	PJTSAU, Telangana
8	CM145	WT	ICAR-IIMR, Ludhiana
9	CM212	WT	ICAR-IIMR, Ludhiana
10	LM11	WT	PAU, Ludhiana
11	LM13	WT	PAU, Ludhiana
12	LM14	WT	PAU, Ludhiana
13	LM17	WT	PAU, Ludhiana
14	E40(3)	WT	MPUAT, Udaipur
15	E128(1)	WT	MPUAT, Udaipur
16	BKI-4(3)	WT	ICAR-IIMR, New Delhi
17	BK9-2	WT	ICAR-IIMR, New Delhi
18	HKI1040-7	WT	CCSHAU, Hisar
19	HKI1344	WT	CCSHAU, Hisar

20	HKI1378	WT	CCSHAU, Hisar
21	PMI-PV1	WT	ICAR-IARI, New Delhi
22	PMI-PV2	WT	ICAR-IARI, New Delhi
23	PMI-PV10	WT	ICAR-IARI, New Delhi
24	PMI-PV12	WT	ICAR-IARI, New Delhi
25	CML51	WT	CIMMYT, Mexico
26	CML116	WT	CIMMYT, Mexico
27	CML150	WT	CIMMYT, Mexico
28	CML165	WT	CIMMYT, Mexico
29	CML202	WT	CIMMYT, Mexico
30	CML320	WT	CIMMYT, Mexico
31	CML385	WT	CIMMYT, Mexico
32	CML410	WT	CIMMYT, Mexico
33	CML451	WT	CIMMYT, Mexico
34	CML465	WT	CIMMYT, Mexico
35	CML470	WT	CIMMYT, Mexico
36	CML473	WT	CIMMYT, Mexico
37	CML474	WT	CIMMYT, Mexico
38	CML487	WT	CIMMYT, Mexico
39	CML511	WT	CIMMYT, Mexico
40	CML550	WT	CIMMYT, Mexico
41	CML559	WT	CIMMYT, Mexico
42	CML602	WT	CIMMYT, Mexico
43	PMI-LP1-110	MT	ICAR-IARI, New Delhi
44	PMI-LP1-112	MT	ICAR-IARI, New Delhi
45	PMI-LP1-115	MT	ICAR-IARI, New Delhi
46	PMI-LP1-116	MT	ICAR-IARI, New Delhi
47	PMI-LP1-120	MT	ICAR-IARI, New Delhi
48	PMI-LP1-123	MT	ICAR-IARI, New Delhi

Table S2. Details of overlapping primers developed for sequencing full length *Lpal-1* gene

S. No.	Primer	Sequence
1	MGU- <i>lpal</i> -A-F	ACCCCAATGCCTTTTAAACC
	MGU- <i>lpal</i> -A-R	GGCCACCTCGTAGCTCAG
2	MGU- <i>lpal</i> -B-F	CGTGCTGAGCTACGAGGTG
	MGU- <i>lpal</i> -B-R	CCATCCTCCGTAAACTCCAA
3	MGU- <i>lpal</i> -C-F	TTGGAGTTTACGGAGGATGG
	MGU- <i>lpal</i> -C-R	CAACGGACTAAGCCATGACA
4	MGU- <i>lpal</i> -D-F	TTGCTGCTGTCAACACGATT
	MGU- <i>lpal</i> -D-R	ATGGCGAGAGCAAGAATGAT
5	MGU- <i>lpal</i> -E-F	AAAGCTGCAGGAGCACTACC
	MGU- <i>lpal</i> -E-R	ATGCGCCATCCTTAATATCG
6	MGU- <i>lpal</i> -F-F	TGGAACCCATACACTCTGACC
	MGU- <i>lpal</i> -F-R	CCAGTATGAGCATCAACAGCA
7	MGU- <i>lpal</i> -G-F	TCCCTTCAGTGCTGTTGATG
	MGU- <i>lpal</i> -G-R	TTTGTAAGCTTCCCCCATGT
8	MGU- <i>lpal</i> -H-F	AAGGGAACGTGGAAAAGTGA
	MGU- <i>lpal</i> -H-R	ACCAGATGGTGTGGTGTCAA
9	MGU- <i>lpal</i> -I-F	GCGTTCAGACTTGGTGGATT
	MGU- <i>lpal</i> -I-R	CATGCAAAAAGCAAAGACGA
10	MGU- <i>lpal</i> -J-F	GGAATTGCTTTCGACTTTTCG
	MGU- <i>lpal</i> -J-R	CCGCCAGGAAACATACAACCT
11	MGU- <i>lpal</i> -K-F	TCAAGGTCCGCTACAAGGAC
	MGU- <i>lpal</i> -K-R	TTGGAACGAATGACCTCTCC
12	MGU- <i>lpal</i> -L-F	GGAGAGGTCATTCGTTCCAA
	MGU- <i>lpal</i> -L-R	GTGGGAATACGGTGAGCAAT
13	MGU- <i>lpal</i> -M-F	CCGTATTCCCACCGTTATTG
	MGU- <i>lpal</i> -M-R	CATGGCAGATACGGGCTATT

F-Forward, R-Reverse

Table S3. Primer details employed for gene-based diversity in *Lpa1-1* gene

S. no	Marker	Sequence (5'→3')	Amplicon size (bp)	Region
1	<i>lpa1</i> -InDel-1F	GGATTACGCGCACATGGT	72	E1
	<i>lpa1</i> -InDel-1R	ATGACACCAACCAAGCACAG		
2	<i>lpa1</i> -InDel-2F	GATCCTCAGCCTTGCAACAT	86	E1
	<i>lpa1</i> -InDel-2R	AGGGTATGTCAGCCAACTCAA		
3	<i>lpa1</i> -InDel-3F	GACCGTGCAAAGTCATGCTA	72	E1
	<i>lpa1</i> -InDel-3R	CCTGCCAGGGTATTCTAGCC		
4	<i>lpa1</i> -InDel-4F	TTGCTGCTGTCAACACGATT	75	E1
	<i>lpa1</i> -InDel-4R	TGCCACTGAGGTAGTCCACA		
5	<i>lpa1</i> -InDel-5F	TACTCGGGGAGATACCCAAA	158	E1-E2
	<i>lpa1</i> -InDel-5R	TCCATGCAGTCTGAGGAACA		
6	<i>lpa1</i> -InDel-6F	TCCCTTCAGTGCTGTTGATG	154	E2-E3
	<i>lpa1</i> -InDel-6R	TTTTGGTTGCTAGTGCAGTCA		
7	<i>lpa1</i> -InDel-7F	TGGCTACGTTTGGTTTAGCA	100	E4
	<i>lpa1</i> -InDel-7R	CCAGATGGTGTGGTGTCAAA		
8	<i>lpa1</i> -InDel-8F	GCGTTCAGACTTGGTGGATT	80	E5
	<i>lpa1</i> -InDel-8R	TGCCATGTGACTTTGCTCAT		
9	<i>lpa1</i> -InDel-9F	CCTTGGGAATTGTTGCTGTCA	575	E5-E6
	<i>lpa1</i> -InDel-9R	AATCCTAGTTAGTTCCCTTGATGA		
10	<i>lpa1</i> -InDel-10F	CTCTGCCTGCGAATGGAAT	101	E6-I6
	<i>lpa1</i> -InDel-10R	ATAAACATACTTGGTTTCGAT		
11	<i>lpa1</i> -InDel-11F	CATGGCCTCAGAATGGAAAC	308	E7-E8
	<i>lpa1</i> -InDel-11R	CCGCCAGGAAACATACTAACT		
12	<i>lpa1</i> -InDel-12F	GAATGAACCTTGATCCTCTTGA	159	E8-E9
	<i>lpa1</i> -InDel-12R	TGACCTCTCCTAGCTGACACTTT		
13	<i>lpa1</i> -InDel-13F	CCGTATTCCCACCGTTATTG	278	E10-E11
	<i>lpa1</i> -InDel-13R	CGTGTCGAACTCTGCGATT		
14	<i>lpa1</i> -InDel-14F	CCCAAACCTGGCAACAGA	75	E11
	<i>lpa1</i> -InDel-14R	TTCGCAGGGACCTCTATGG		
15	<i>lpa1</i> -InDel-15F	ACATTACCCGAAGCTGATGC	85	E11
	<i>lpa1</i> -InDel-15R	TCCCTTGGTTCTCCACTCTG		

F represents Forward; R represents Reverse; E represents exon

Table S4. Evolutionary distance and synonymous and non-synonymous scores in comparison with reference *Lpal* (Gene ID: Zm00001eb003490)

S. No.	Inbred	Pairwise distance	Ks	Ka	Ka/Ks
1	<i>Lpal</i> -wild-1	0.00922	0.0073	0.0051	0.69863
2	<i>Lpal</i> -wild-2	0.00962	0.0085	0.0057	0.670588
3	<i>Lpal</i> -wild-3	0.01346	0.0105	0.0087	0.828571
4	<i>Lpal</i> -wild-4	0.01264	0.0098	0.009	0.918367
5	<i>Lpal</i> -wild-5	0.01875	0.0149	0.0158	1.060403
6	<i>Lpal</i> -wild-6	0.01306	0.0095	0.0079	0.831579
7	<i>Lpal</i> -wild-7	0.01209	0.0129	0.009	0.697674
8	<i>lpal-l</i> -mutant-1	0.01237	0.0141	0.0111	0.787234
9	<i>lpal-l</i> -mutant-2	0.01894	0.0178	0.0141	0.792135
10	<i>lpal-l</i> -mutant-3	0.00603	0.0063	0.0044	0.698413

Table S5. Structural characteristics of the *Lpal-1* gene in maize genotypes and selected group of orthologue accessions

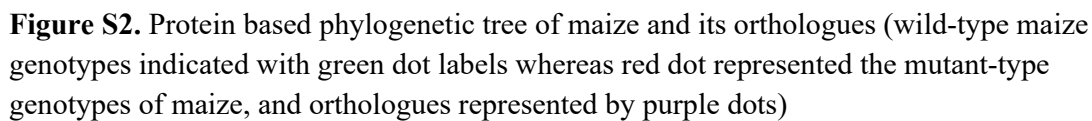
S. No.	Accession	Crop species	Family	No. of exons	Exon range (bp)	Intron range (bp)	Coding sequence
1	Zm00001eb003490	<i>Zea mays</i>	Poaceae	11	66-2314	67-445	4530
2	<i>Lpal</i> -wild-1		Poaceae	19	21-489	38-639	3588
3	<i>Lpal</i> -wild-2		Poaceae	14	63-981	76-575	3381
4	<i>Lpal</i> -wild-3		Poaceae	16	21-576	36-450	3228
5	<i>Lpal</i> -wild-4		Poaceae	14	63-984	49-537	3825
6	<i>Lpal</i> -wild-5		Poaceae	16	48-771	52-659	3255
7	<i>Lpal</i> -wild-6		Poaceae	16	39-1059	75-450	3558
8	<i>Lpal</i> -wild-7		Poaceae	13	63-639	79-450	3528
9	<i>lpal-1</i> -mutant-1		Poaceae	14	21-597	66-892	3012
10	<i>lpal-1</i> -mutant-2		Poaceae	14	21-378	76-483	1896
11	<i>lpal-1</i> -mutant-3		Poaceae	17	21-552	76-635	3402
12	AET4Gv20803900	<i>Aegilops tauschii</i>	Poaceae	11	63-1782	55-502	4365
13	Aco010163	<i>Ananas comosus</i>	Bromeliaceae	9	63-2076	93-947	4125
14	A4U43_C10F18740	<i>Asparagus officinalis</i>	Asparagaceae	6	6-219	58-101	387
15	BRADI_1g75590v3	<i>Brachypodium distachyon</i>	Poaceae	10	63-2046	77-805	3933
16	DRNTG_05198	<i>Dioscorea rotundata</i>	Dioscoreaceae	11	63-2043	87-1350	4566
17	EJB05_08061	<i>Eragrostis curvula</i>	Poaceae	12	63-1944	95-635	4584
18	HORVU.MOREX.r3.4HG0412040	<i>Hordeum vulgare</i>	Poaceae	11	63-2049	80-564	4281
19	LPERR03G03070	<i>Leersia perrieri</i>	Poaceae	11	39-2043	76-637	3666
20	Ma08_g12530	<i>Musa acuminata</i>	Musaceae	11	63-2046	74-2449	4590
21	Ma11_g02290	<i>Musa acuminata</i>	Musaceae	13	42-2043	71-2653	4626
22	OBART03G03330	<i>Oryza barthii</i>	Poaceae	10	63-1683	80-450	3711

23	OB03G13350	<i>Oryza brachyantha</i>	Poaceae	10	60-1671	81-769	3384
24	GQ55_9G618800	<i>Panicum hallii HAL2</i>	Poaceae	7	63-2040	78-2074	2934
25	SECCE5Rv1G0369730	<i>Secale cereale</i>	Poaceae	12	63-2058	80-372	4704
26	SEVIR_9G548400v2	<i>Setaria viridis</i>	Poaceae	14	63-2049	76-431	5226
27	SORBI_3001G508200	<i>Sorghum bicolor</i>	Poaceae	9	63-2067	77-1019	3342
28	TraesCS5A02G512500	<i>Triticum aestivum</i>	Poaceae	11	63-1914	79-374	4383
29	TRIDC4BG057910	<i>Triticum dicoccoides</i>	Poaceae	11	63-1911	79-414	4215
30	TRITD5Av1G244640	<i>Triticum turgidum</i>	Poaceae	11	63-1914	79-374	4383
31	TuG1812G0500005238.01	<i>Triticum urartu</i>	Poaceae	11	63-2061	79-374	4530
32	OsABCC13	<i>Oryza sativa Japonica Group</i>	Poaceae	12	63-2043	76-705	4341
33	ABCC13 BGIOGA011835	<i>Oryza sativa Indica Group</i>	Poaceae	10	63-2043	76-450	4071
34	ORUF103G03050	<i>Oryza rufipogon</i>	Poaceae	12	63-2046	76-705	4344
35	ONIVA03G03060	<i>Oryza nivara</i>	Poaceae	11	63-2043	76-450	4233
36	Et_s2678-1.30-1.path1	<i>Eragrostis tef</i>	Poaceae	11	27-537	79-384	2133
37	SETIT_033885mg	<i>Setaria italica</i>	Poaceae	12	63-2049	77-431	4305

Table S6. *Zea mays* LPA1 protein (Protein ID: A7KVC2) domains and features

S. No.	Domain source	Start	End	Description	Accession
1	PANTHER	83	1510	-	PTHR24223
2	PANTHER	83	1510	-	PTHR24223:SF391
3	Gene3D	289	612	ABC transporter type 1, transmembrane domain superfamily	1.20.1560.10
4	Superfamily	303	613	ABC transporter type 1, transmembrane domain superfamily	SSF90123
5	CDD	316	603	ABC transporter C family, six-transmembrane helical domain 1	cd18579
6	Pfam	319	583	ABC transporter type 1, transmembrane domain	PF00664
7	PROSITE profiles	320	595	ABC transporter type 1, transmembrane domain	PS50929

8	Superfamily	625	851	P-loop containing nucleoside triphosphate hydrolase	SSF52540
9	Gene3D	627	859	P-loop containing nucleoside triphosphate hydrolase	3.40.50.300
10	PROSITE profiles	629	852	ABC transporter-like, ATP-binding domain	PS50893
11	CDD	629	828	-	cd03250
12	Pfam	647	781	ABC transporter-like, ATP-binding domain	PF00005
13	SMART	656	829	AAA+ ATPase domain	SM00382
14	PROSITE patterns	754	768	ABC transporter-like, conserved site	PS00211
15	Gene3D	908	1248	ABC transporter type 1, transmembrane domain superfamily	1.20.1560.10
16	Superfamily	947	1249	ABC transporter type 1, transmembrane domain superfamily	SSF90123
17	CDD	948	1242	ABC transporter C family, six-transmembrane helical domain 2	cd18580
18	PROSITE profiles	950	1220	ABC transporter type 1, transmembrane domain	PS50929
19	Pfam	951	1190	ABC transporter type 1, transmembrane domain	PF00664
20	Gene3D	1250	1502	P-loop containing nucleoside triphosphate hydrolase	3.40.50.300
21	Superfamily	1259	1499	P-loop containing nucleoside triphosphate hydrolase	SSF52540
22	CDD	1265	1485	-	cd03244
23	PROSITE profiles	1267	1501	ABC transporter-like, ATP-binding domain	PS50893
24	Pfam	1284	1432	ABC transporter-like, ATP-binding domain	PF00005
25	SMART	1293	1478	AAA+ ATPase domain	SM00382



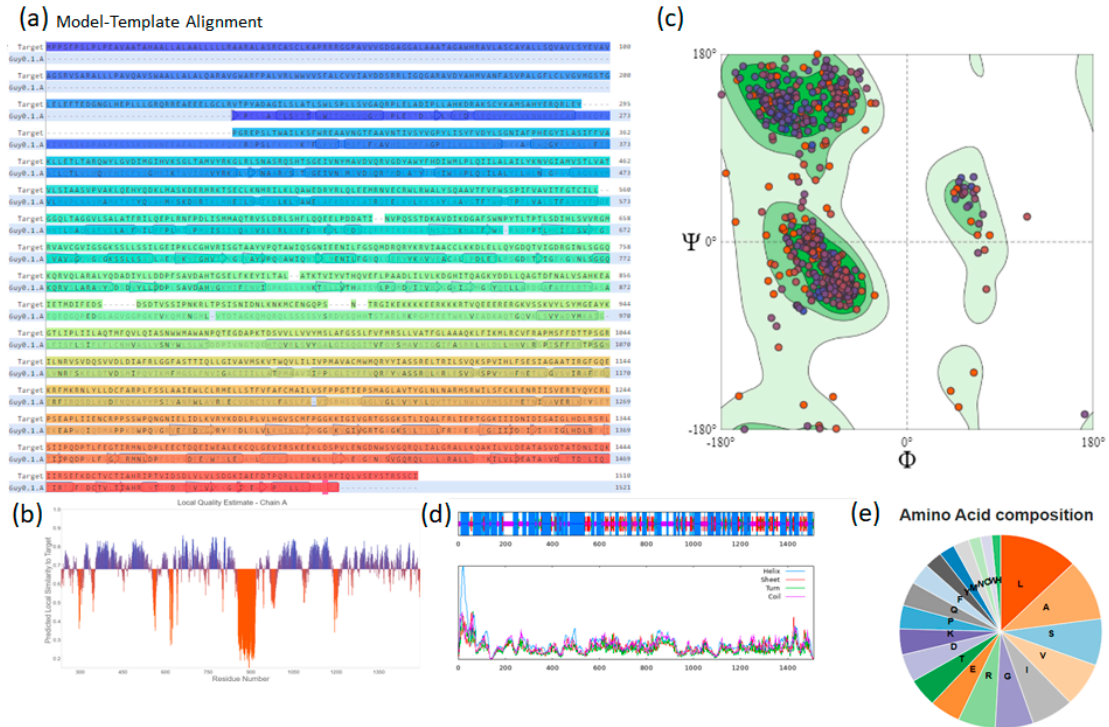


Figure S3. Homology based LPA1 protein (protein ID: A7KVC2), (a) alignment with 6uy0.1.A (Bovine) template by Swiss Model; (b) Local quality estimation (c) Ramachandran plot using Swiss Model (d) Secondary structure by SOPMA (e) amino acid compositions by predict protein tool