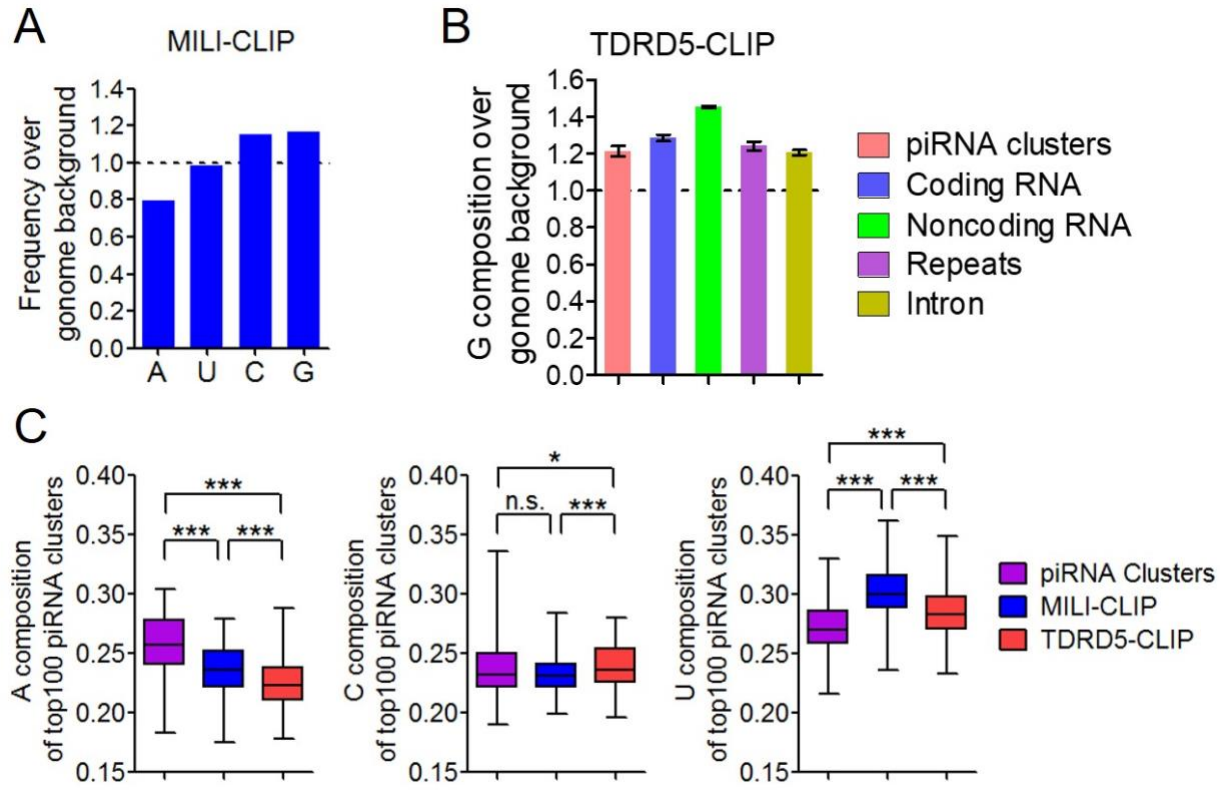


Figure S1. TDRD5 binds to G-rich RNA *in vivo*. (**A**) Nucleotide composition of MILI CLIP reads was normalized to that of the mouse genome. (**B**) TDRD5-CLIP reads contain higher G-contents over the mouse genome background. TDRD5-CLIP reads were mapped to the mouse genome. G-content of each genomic region was normalized to that of the mouse genome. (**C**) Nucleotide composition of top 100 piRNA clusters and MILI-CLIP or TDRD5-CLIP reads that are mapped to top100 piRNA clusters were shown. n.s., Nonsignificant; *, $p < 0.05$; ***, $p < 0.001$.

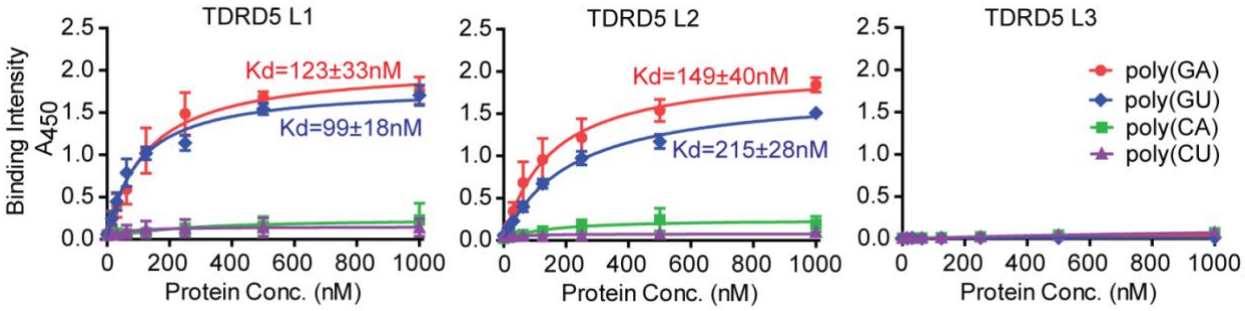


Figure S2. LOTUS domains bind to G-rich RNA by ELISA assays. n=3, error bars represent s.e.m. Dissociation constants (Kd) are indicated.

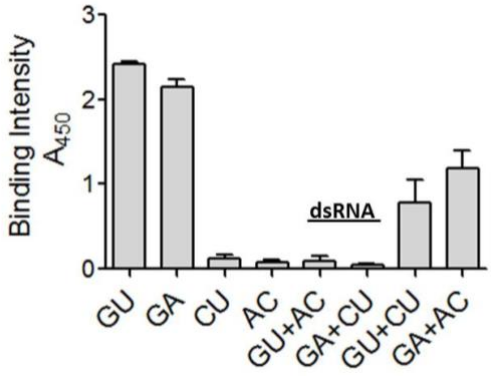


Figure S3. LOTUS domain binds G-rich ssRNA, not dsRNA. The indicated biotin-labelled oligonucleotides were annealed. ELISA assays were performed using His-tagged TDRD5 L1 and the annealed RNAs. The mixtures of poly(GU) -poly(AC) and poly(GA)-poly(CU) form the dsRNA.

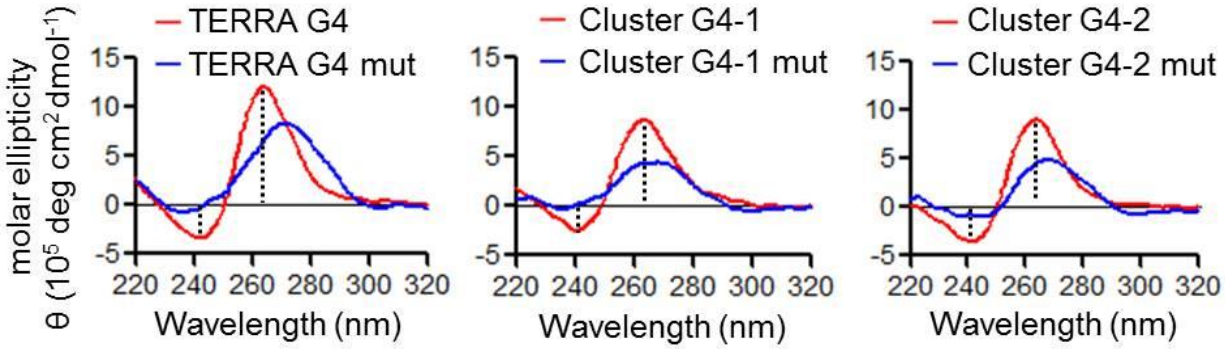


Figure S4. Circular dichroism spectroscopy of RNA G4 and their mutant oligonucleotides. TERRA G4, Cluster G4-1, and Cluster G4-2 RNA oligonucleotides show one positive peak at ~260 nm and one negative peak at ~240 nm, the characteristic of parallel G4 structure. On the contrary, their mutant oligonucleotides do not show G4 secondary structure signature.

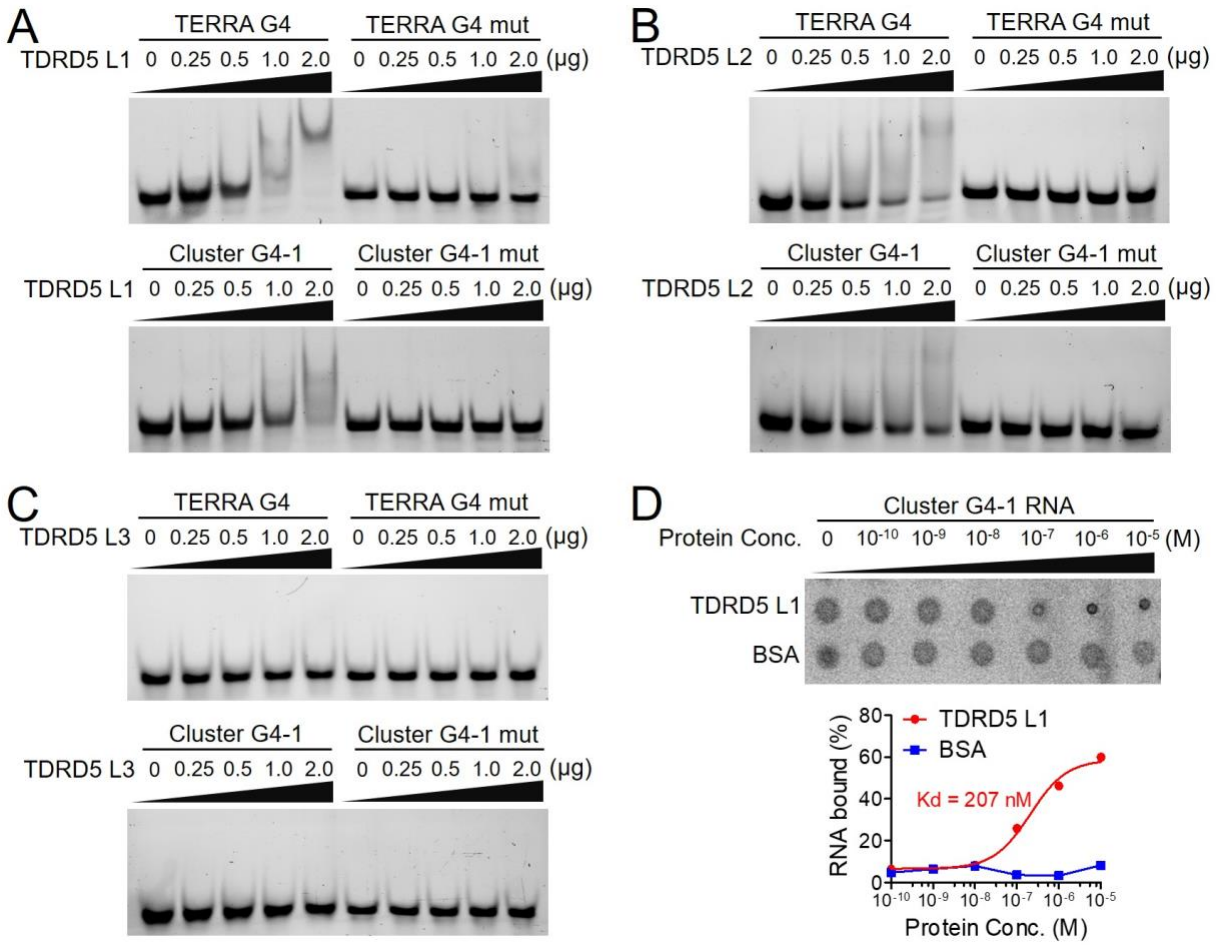
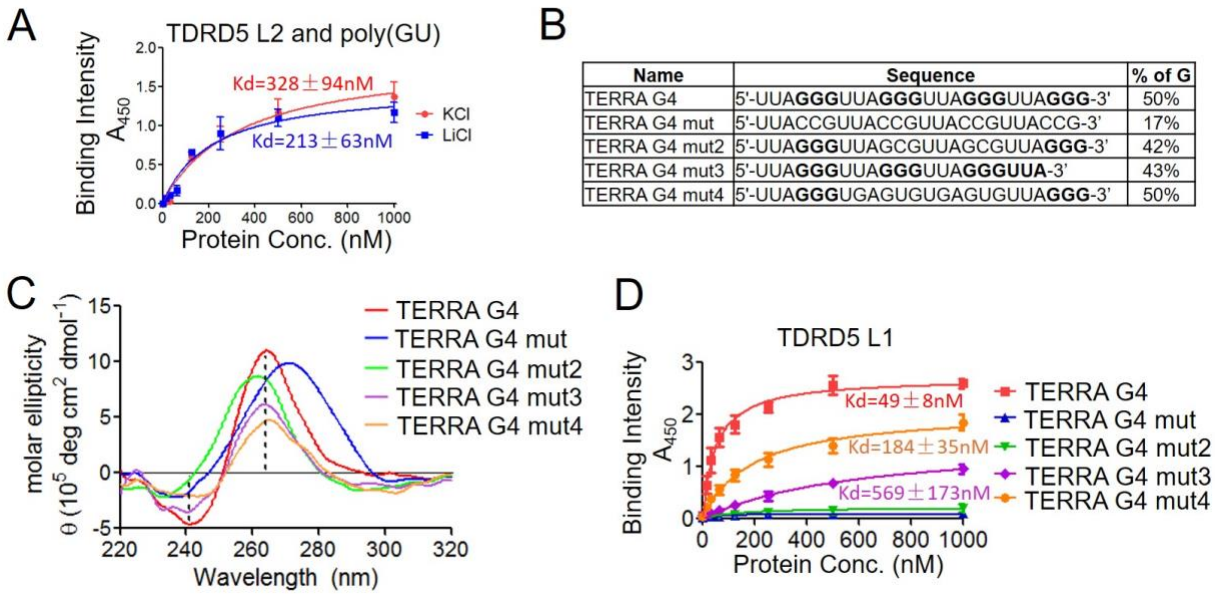


Figure S5. TDRD5 LOTUS domains bind to RNA G4. (**A**) TDRD5 L1 binds to G4 by EMSA assay. 5-FAM-labelled RNA G4 oligonucleotides and their respective mutants (mut) were incubated with purified His-tagged TDRD5 L1. Samples were separated by native PAGE gel electrophoresis and fluorescent signals were captured. (**B**) TDRD5 L2 binds to G4 by EMSA assay. (**C**) TDRD5 L3 does not bind to G4 by EMSA assay. (**D**) DRaCALA protein-RNA assay showing specific LOTUS domain-G4 interaction. ^{32}P-labelled Cluster G4-1 RNA was incubated with TDRD5 L1 at indicated concentrations. Bovine serum albumin (BSA) was used as a negative control for G4 binding. Dissociation constants (Kd) are indicated.



Name	Sequence	% of G
TERRA G4	5'-UUA**GGG**UUA**GGG**UUA**GGG**UUA**GGG**-3'	50%
TERRA G4 mut	5'-UUACCGUUACCGUUACCGUUACCG-3'	17%
TERRA G4 mut2	5'-UUA**GGG**UUAGCGUUAGCGUUA**GGG**-3'	42%
TERRA G4 mut3	5'-UUA**GGG**UUA**GGG**UUA**GGGUUA**-3'	43%
TERRA G4 mut4	5'-UUA**GGG**UGAGUGUGAGUGUUA**GGG**-3'	50%

Figure S6. LOTUS domains recognize G4 tertiary structure. (**A**) ELISA assay was performed with TDRD5 L2 and poly(GU) RNA oligos in KCl or LiCl buffers. n=3, error bars represent s.e.m. Dissociation constants (Kd) are indicated. (**B**) Sequences of TERRA G4 RNA and TERRA G4 mutant RNAs. G composition of each RNA oligo is shown. (**C**) CD spectroscopy of TERRA G4 and TERRA G4 mutants. (**D**) ELISA assay was performed to measure the interaction of TDRD5 L1 with TERRA G4 and 4 different TERRA G4 mutants. n=3, error bars represent s.e.m. Dissociation constants (Kd) are indicated.

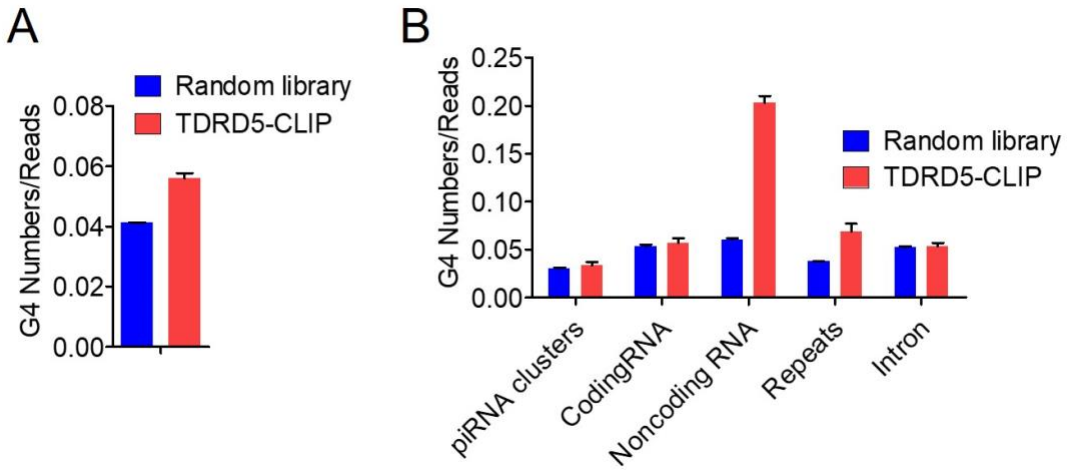


Figure S7. TDRD5-CLIP reads contain higher frequency of G4 forming sequences compared to the mouse genome. (**A**) RNA G4 frequency in TDRD5-CLIP reads and random reads from the mouse genome. (**B**) RNA G4 frequency of TDRD5-CLIP reads in different genomic regions. Random reads from each genomic region were used to calculate the control baseline RNA G4 frequency.

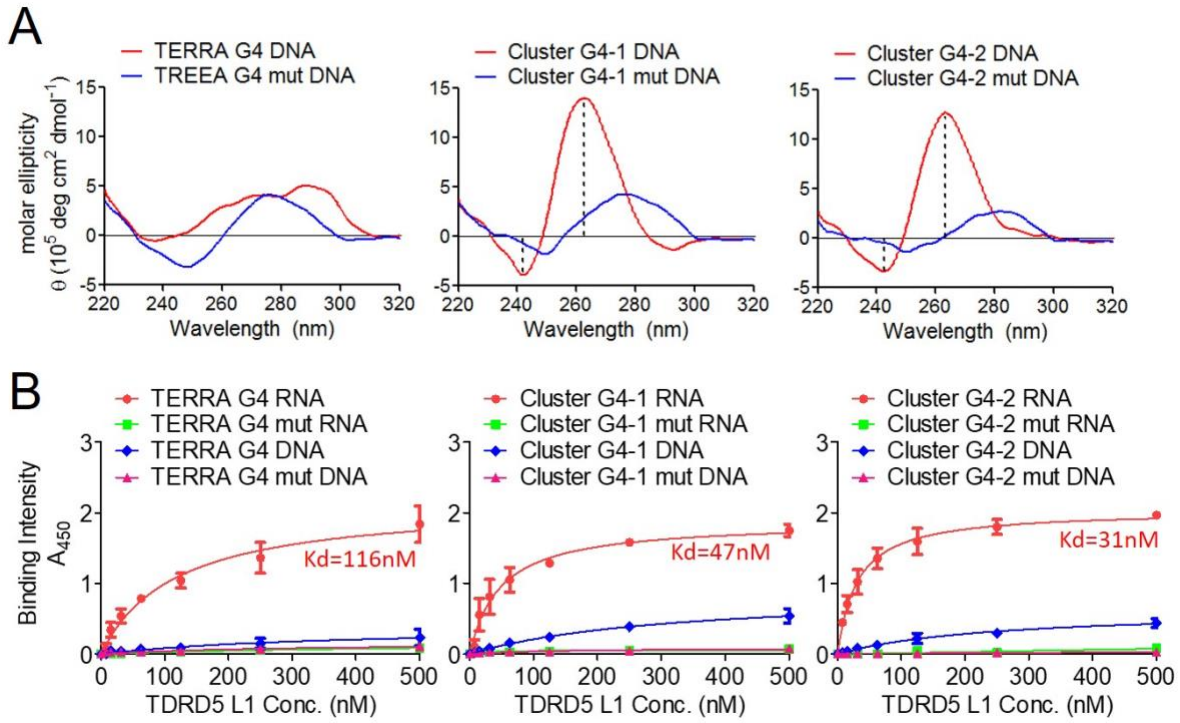


Figure S8. LOTUS domain preferentially recognizes RNA G4 but not DNA G4. (**A**) CD spectroscopy of DNA G4 and G4 mutant oligonucleotides. Cluster G4-1 DNA and Cluster G4-2 DNA oligonucleotides form parallel G4 structure, while TERRA G4 DNA forms anti-parallel G4. On the contrary, their mut oligonucleotides do not show G4 signature. (**B**) ELISA assay was performed to determine the binding of TDRD5 L1 to the indicated DNA or RNA oligonucleotides. n=2, error bars represent s.e.m. Dissociation constants (Kd) are indicated.

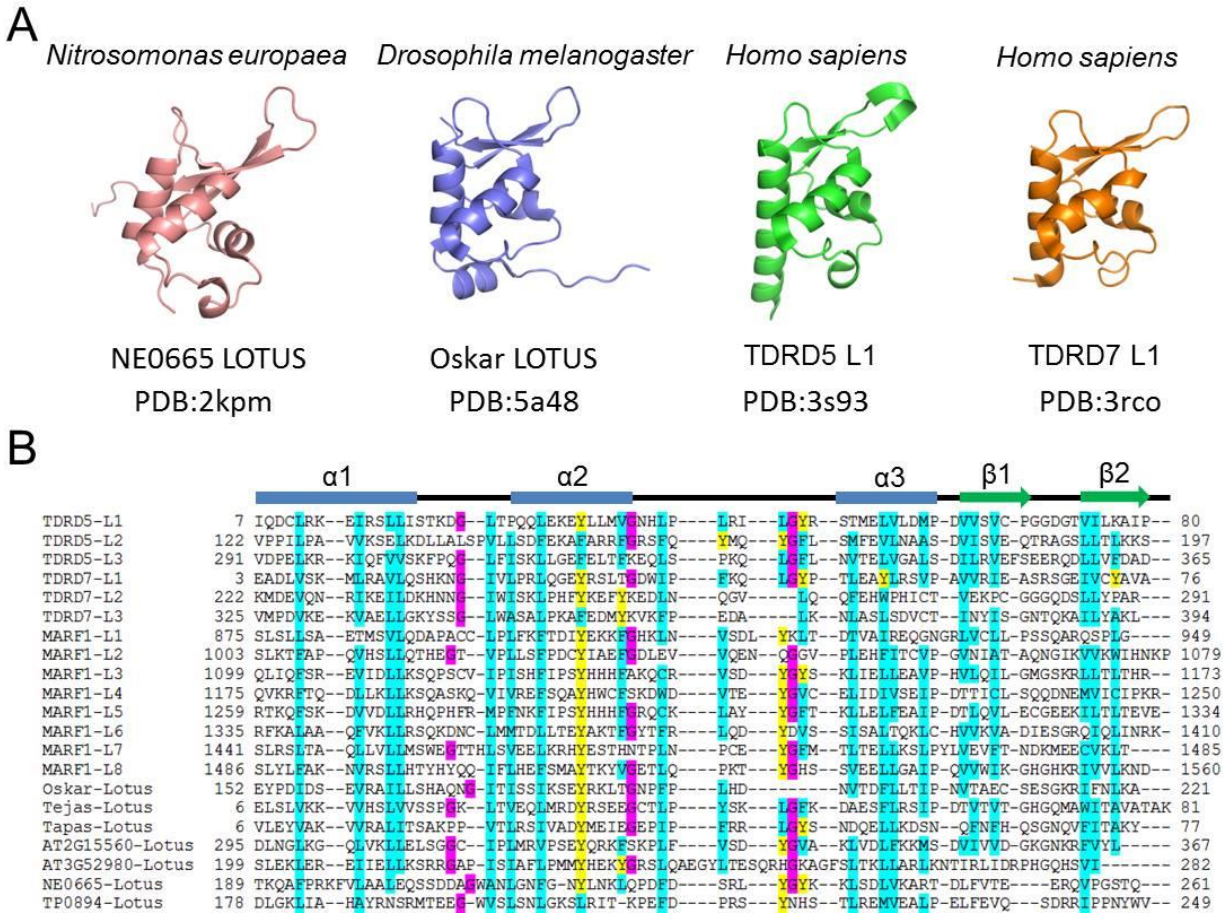


Figure S9. The structural fold of LOTUS domains is conserved. (**A**) Crystal structures of LOTUS domains deposited in Protein Data Bank (PDB). (**B**) Multiple sequence alignment of LOTUS domains from diverse proteins of different species. Shown are LOTUS domains from: *Mus musculus* TDRD5 and TDRD7, *Homo sapiens* MARF1, *Drosophila melanogaster* Oskar, Tejas and Tapas, *Arabidopsis thaliana* AT2G15560 and AT3G52980, *Nitrosomonas europaea* NE0665, *Treponema pallidum subsp. pallidum str. Nichols* TP0894. The LOTUS domain sequences were obtained from UniProtKB. Sequence alignment was generated using Clustal X. The conserved hydrophobic (cyan), glycine (violet) and tyrosine (yellow) residuals were highlighted, respectively. The secondary structures of LOTUS domains were predicted using JPred4 (http://www.compbio.dundee.ac.uk/jpred/). Conserved three α-helices (α1, α2 and α3) and two β-sheets (β1 and β2) of LOTUS domains are shown.

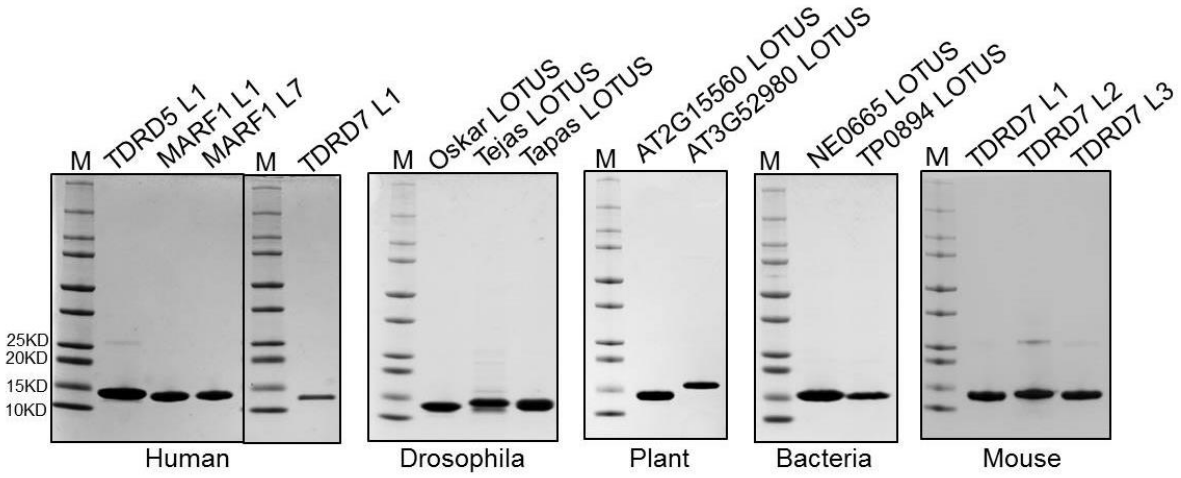


Figure S10. Purification of LOTUS domains derived from various species. His-tagged LOTUS domains were purified by affinity chromatography followed by gel filtration chromatography. Coomassie Brilliant Blue staining was performed after SDS-PAGE electrophoresis.

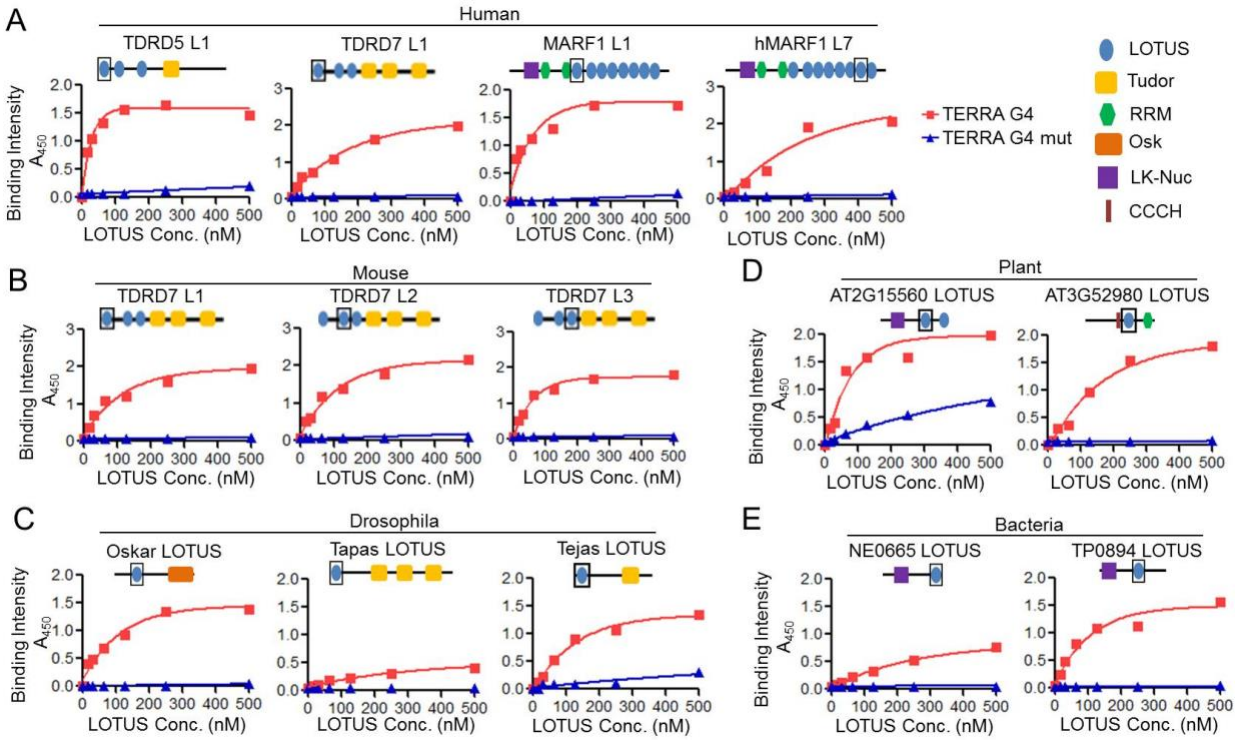


Figure S11. LOTUS domain-RNA G4 interaction is evolutionarily conserved. (**A-E)** Binding of representative LOTUS domains from diverse species to RNA G4 (TERRA G4 and its mutant) was determined by ELISA assay. A representative graph of 2-3 technical repeats is shown for each LOTUS domain. (**A**) Human LOTUS domains. (**B**) Mouse LOTUS domains. (**C**) *Drosophila* LOTUS domains. (**D**) Plant LOTUS domains. (**E**) Bacterial LOTUS domains. His-tagged LOTUS domains were used to bind immobilized biotin-G4 or its G4 mutant in the ELISA assay. The concentration (conc.) of LOTUS domains used is shown on the x-axis of each graph. The position of each LOTUS domain used is shown as a box in cartoons of the protein architecture.

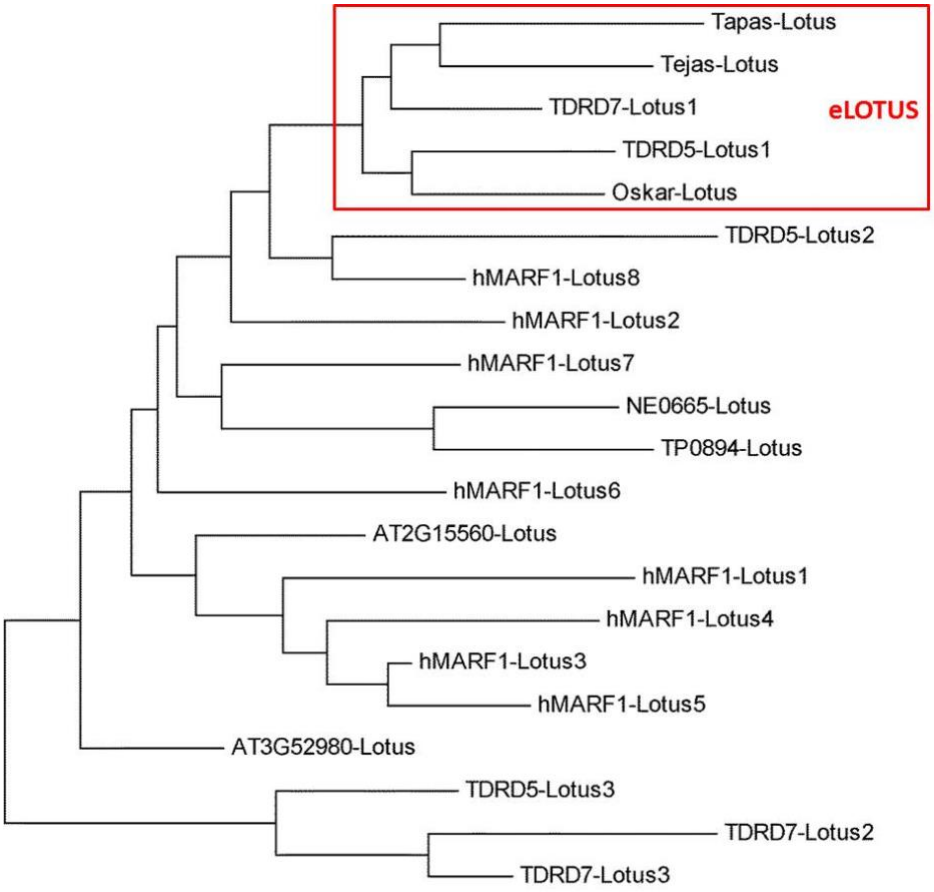


Figure S12. Phylogenetic analysis of LOTUS domains. The LOTUS domain sequences were obtained from UniProtKB (Supplementary Figure S9). The phylogenic tree was constructed with MEGA5 using the Maximum Likelihood method. eLOTUS domains are highlighted in the red frame.

Table S1. Oligonucleotides used in protein-RNA binding assays.

Biotin-oligo		Sequence(5'-3')	% of G
RNA	Poly(G)	GGGGGGGGGGGGGGGGGGGG	100%
	Poly(C)	CCCCCCCCCCCCCCCCCCCC	0%
	Poly(U)	UUUUUUUUUUUUUUUUUUUU	0%
	Poly(A)	AAAAAAAAAAAAAAAAAAAA	0%
	Poly(GU)	GUGUGUGUGUGUGUGUGUGUGUGUGUGU	50%
	Poly(GA)	GAGAGAGAGAGAGAGAGAGAGAGAGAGA	50%
	Poly(CU)	CUCUCUCUCUCUCUCUCUCUCUCUCUCU	0%
	Poly(CA)	CACACACACACACACACACACACACACA	0%
	TERRA G4	UUAGGGUUAGGGUUAGGGUUAGGG	50%
	TERRA G4 mut	UUACCGUUACCGUUACCGUUACCG	17%
	TERRA G4 mut2	UUAGGGUUAGCGUUAGCGUUAGGG	42%
	TERRA G4 mut3	UUAGGGUUAGGGUUAGGGUUA	43%
	TERRA G4 mut4	UUAGGGUGAGUGUGAGUGUUAGGG	50%
	Cluster G4-1	GGUGGGCAGGGGAGUGGGGGUGGG	75%
	Cluster G4-1 Mut	GGUCCGCAGCCGAGUCCGGGUCCG	42%
	Cluster G4-2	AGGAGGGAAGGGGAGGGGAGGG	73%
	Cluster G4-2 Mut	AGGACCGAAGCCGAGCCGAGCC	36%
DNA	TERRA G4	TTAGGGTTAGGGTTAGGGTTAGGG	50%
	TERRA G4 mut	TTACCGTTACCGTTACCGTTACCG	17%
	Cluster G4-1	GGTGGGCAGGGGAGTGGGGGTGGG	75%
	Cluster G4-1 Mut	GGTCCGCAGCCGAGTCCGGGTCCG	42%
	Cluster G4-2	AGGAGGGAAGGGGAGGGGAGGG	73%
	Cluster G4-2 Mut	AGGACCGAAGCCGAGCCGAGCC	36%

Table S2. Detailed information of LOTUS domains used in this study.

LOTUS Name	Species	Full Name of Protein	Protein ID	Construct Region in Protein
TDRD5 L1	*Mus musculus*	tudor domain-containing protein 5 isoform 1	NP_001128213.1	1-99
TDRD5 L2				115-203
TDRD5 L3				283-371
TDRD7 L1		tudor domain-containing protein 7 isoform 1	NP_001277404.1	34-129
TDRD7 L2				245-329
TDRD7 L3				348-437
MARF1 L1	*Homo sapiens*	meiosis arrest female protein 1 isoform 2	NP_001171927.1	868-956
MARF1 L7				1405-1492
TDRD5 L1		tudor domain-containing protein 5 isoform 1	NP_001186014.1	1-102
TDRD7 L1		tudor domain-containing protein 7 isoform 1	NP_055105.2	1-86
Tapas LOTUS	*Drosophila melanogaster*	CG8920, isoform D	ACZ94485.1	1-98
Oskar LOTUS (eLOTUS)		oskar, isoform A	NP_731295.1	139-241
Oskar mLOTUS		oskar, isoform A	NP_731295.1	139-225
Tejas LOTUS		tejas	NP_610950.2	1-98
AT2G15560 LOTUS	*Arabidopsis thaliana*	AT2G15560 Putative endonuclease or glycosyl hydrolase	AEC06416.1	278-382
AT3G52980 LOTUS		AT3G52980 RNA recognition motif-containing protein	AEE79021.1	183-294
NE0665 LOTUS	*Nitrosomonas europaea*	NE0665 hypothetical protein	CAD84576.1	174-269
TP0894 LOTUS	*Treponema pallidum*	TP_RS04480 hypothetical protein	WP_010882337.1	160-264