Supporting Information

Yuan and Wessler 10.1073/pnas.1104208108

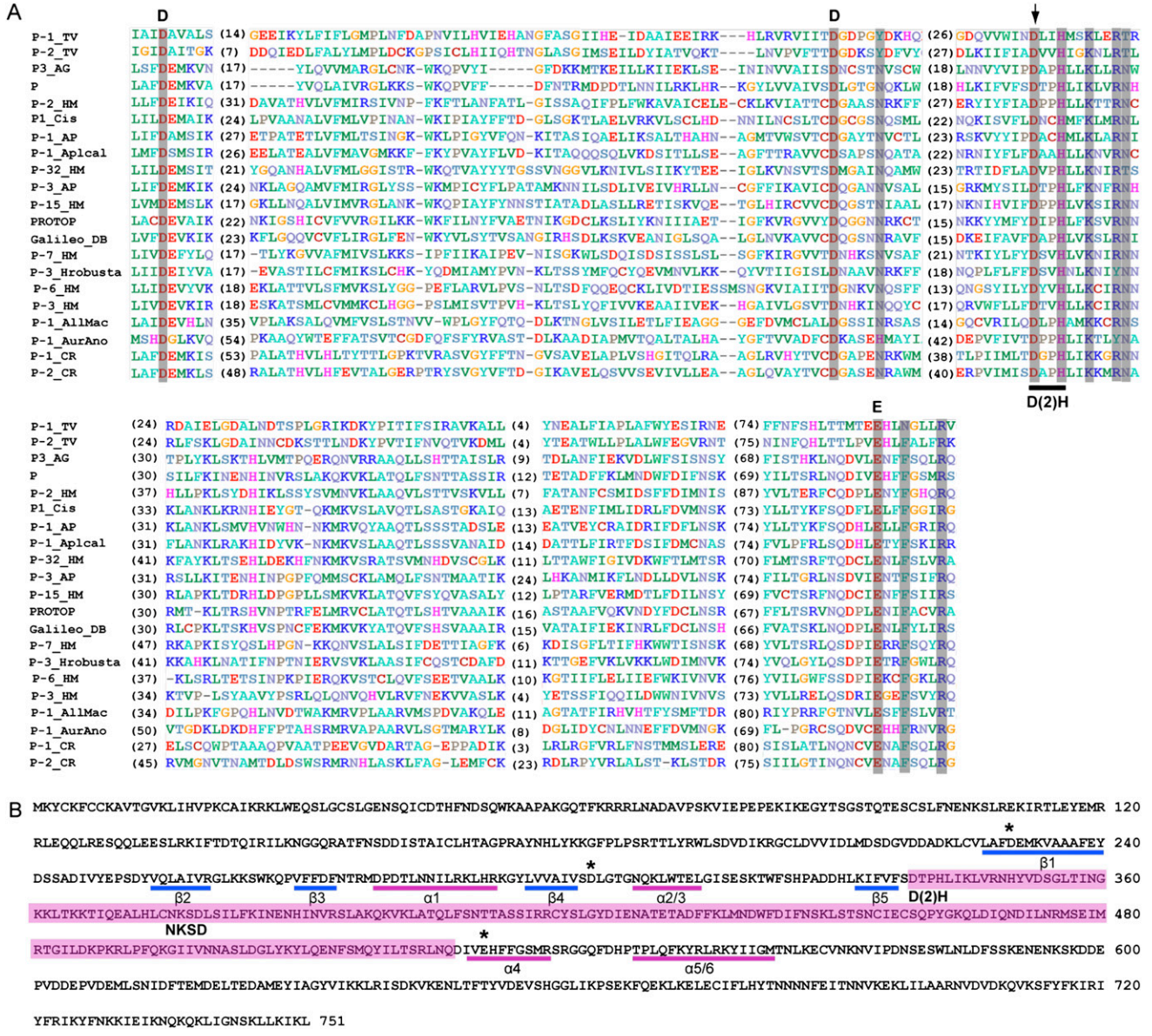

Fig. S1. The DDE domain of *P* elements. (*A*) Alignment shown is after redundancy elimination. Conserved residues are highlighted in gray. The DDE triad, corresponding to D230, D303, and E531 of the *Drosophila melanogaster P* transposase (UniProt: Q7M3K2), are marked above the alignment. Distances between the conserved blocks are indicated in number of amino acid residues. The D(2)H motif is marked below the alignment. The other conserved D in addition to the DDE triad is marked with an arrow. The GTP-binding motif NKSD (1) is located in the variable region after the D(2)H motif. (*B*) Predicted secondary structure of the DDE domain of the *Drosophila melanogaster P* transposase (UniProt: Q7M3K2). The DDE triad is indicated by asterisks. α-Helices and β-strands are highlighted with pink and blue bars, respectively. The first D is on β1, the second D is right after β4, and the E is on α4. The inserted domain between β5 and α4 (highlighted in pink background) is entirely α-helical. The D(2)H motif right after β5 and the NKSD motif in the middle of the inserted domain are also noted.

1. Mul YM, Rio DC (1997) Reprogramming the purine nucleotide cofactor requirement of *Drosophila P* element transposase in vivo. *EMBO J* 16:4441–4447.

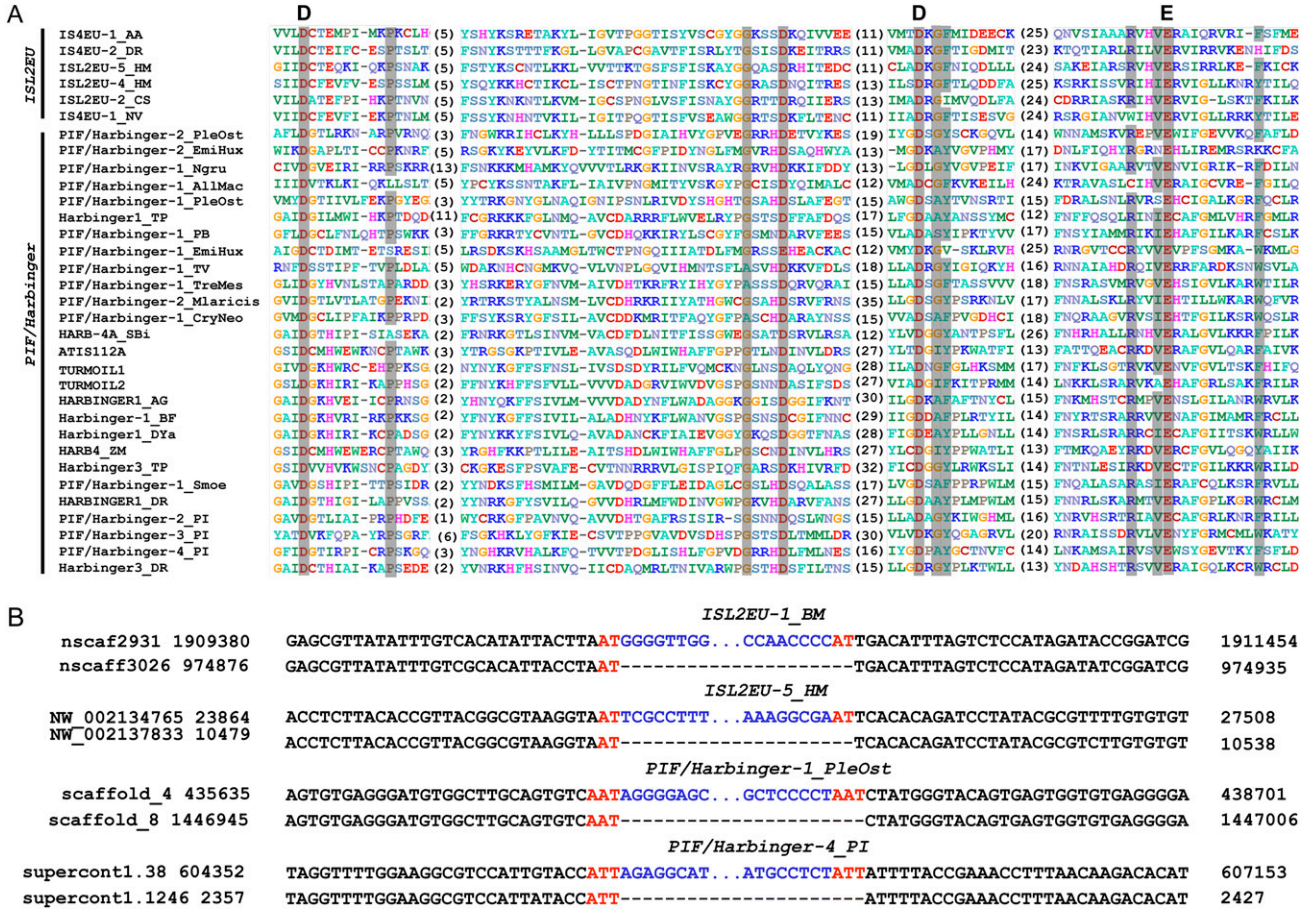

Fig. S2. DDE domains and exemplar TSDs of *PIF/Harbinger* and *ISL2EU* elements. (*A*) Shown is the DDE domain alignment after redundancy elimination. Conserved residues are highlighted in gray. The DDE triad is marked above the alignment. Distances between the conserved blocks are indicated in number of amino acid residues. (*B*) The "AT" target site duplication (TSD) created upon insertion of *ISL2EU* elements and "AWT" TSD created by *PIF/Harbinger* elements from fungi and Chromalveolates protists. Shown are alignments of flanking sequences of the TE insertions with corresponding "empty" paralogous sequences that represent the preinsertion condition. The TSDs are in red, and the 5′- and 3′- terminal nucleotides are in purple. The four elements are from *Bombyx mori* (*ISL2EU-1_BM*), *Hydra magnipapillata* (*ISL2EU-5_HM*), *Pleurotus ostreatus* (*PIF/Harbinger-1_PleOst*), and *Phytophtora infestans* (*PIF/Harbinger-4_PI*).

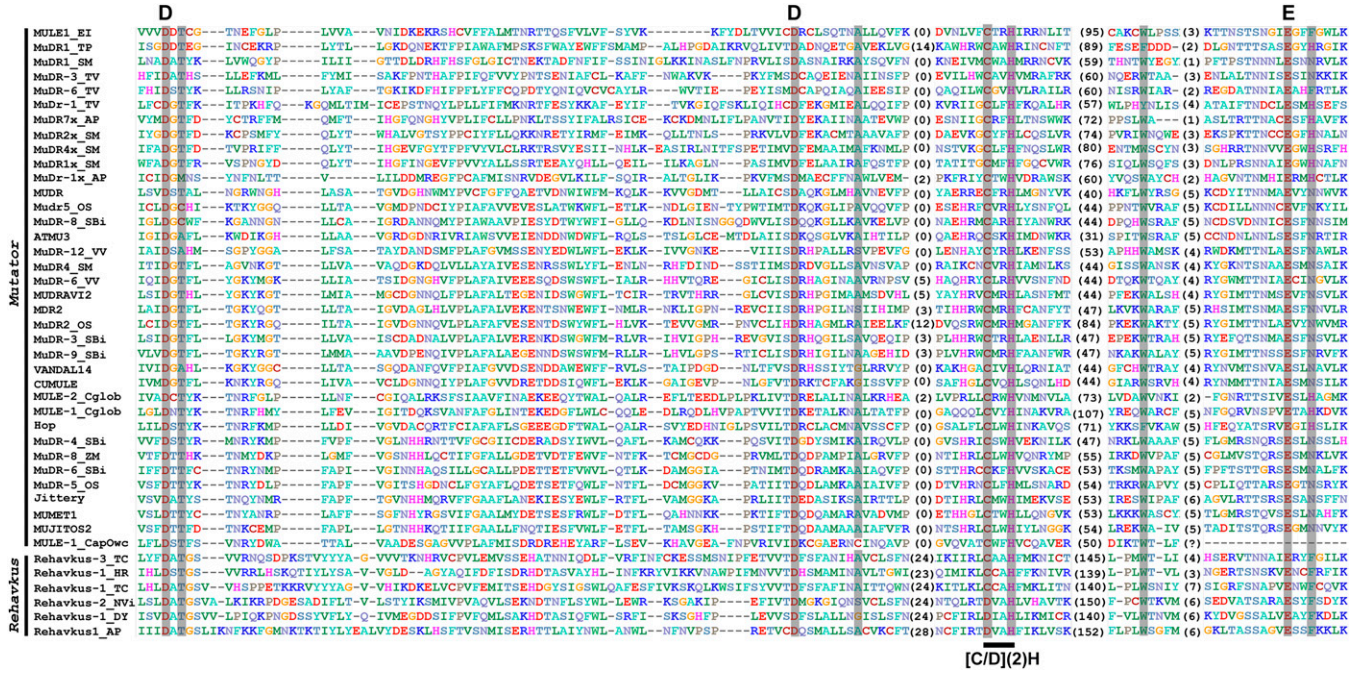

Fig. S3. DDE domains of *Mutator* and *Rehavkus* elements. Alignment shown is after redundancy elimination. Conserved residues are highlighted in gray. The DDE triad is marked above the alignment. The [C/D](2)H motif is marked below the alignment. Distances between the conserved blocks are indicated in number of amino acid residues.

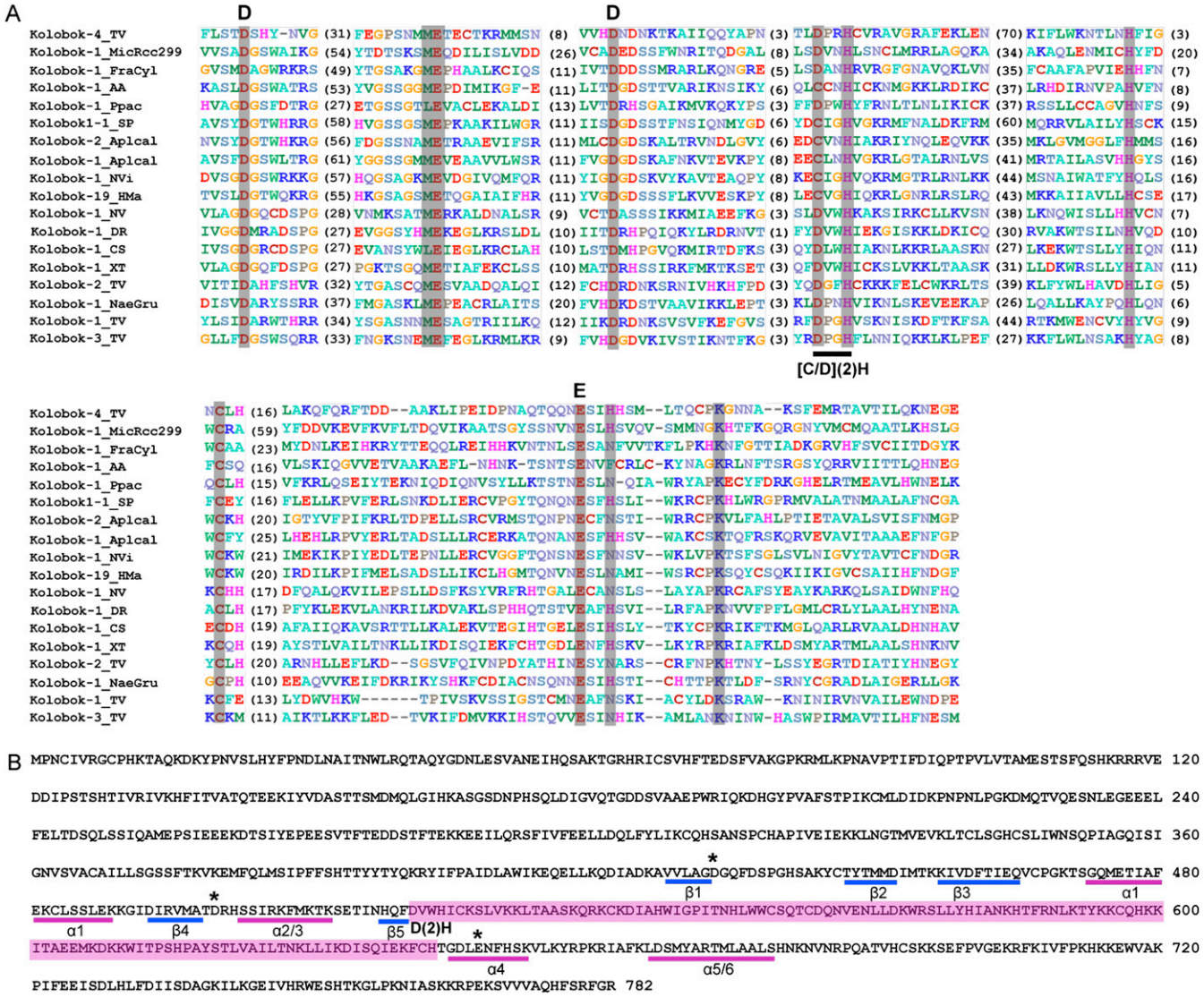

Fig. S4. The DDE domain of *Kolobok* elements. (*A*) Alignment shown is after redundancy elimination. Conserved residues are highlighted in gray. The DDE triad is marked above the alignment. The [C/D](2)H motif is also noted. Distances between the conserved blocks are indicated in number of amino acid residues. (*B*) Predicted secondary structure of the DDE domain of the *Kolobok-1_XT* transposase. The DDE triad is indicated by asterisks. α-Helices and β-strands are highlighted with pink and blue bars, respectively. The first D is associated with β1, the second D is right after β4, and the E is on α4. The inserted domain between β5 and α4 (highlighted in pink) is predominantly α-helical. Also note the D(2)H motif right after β5.

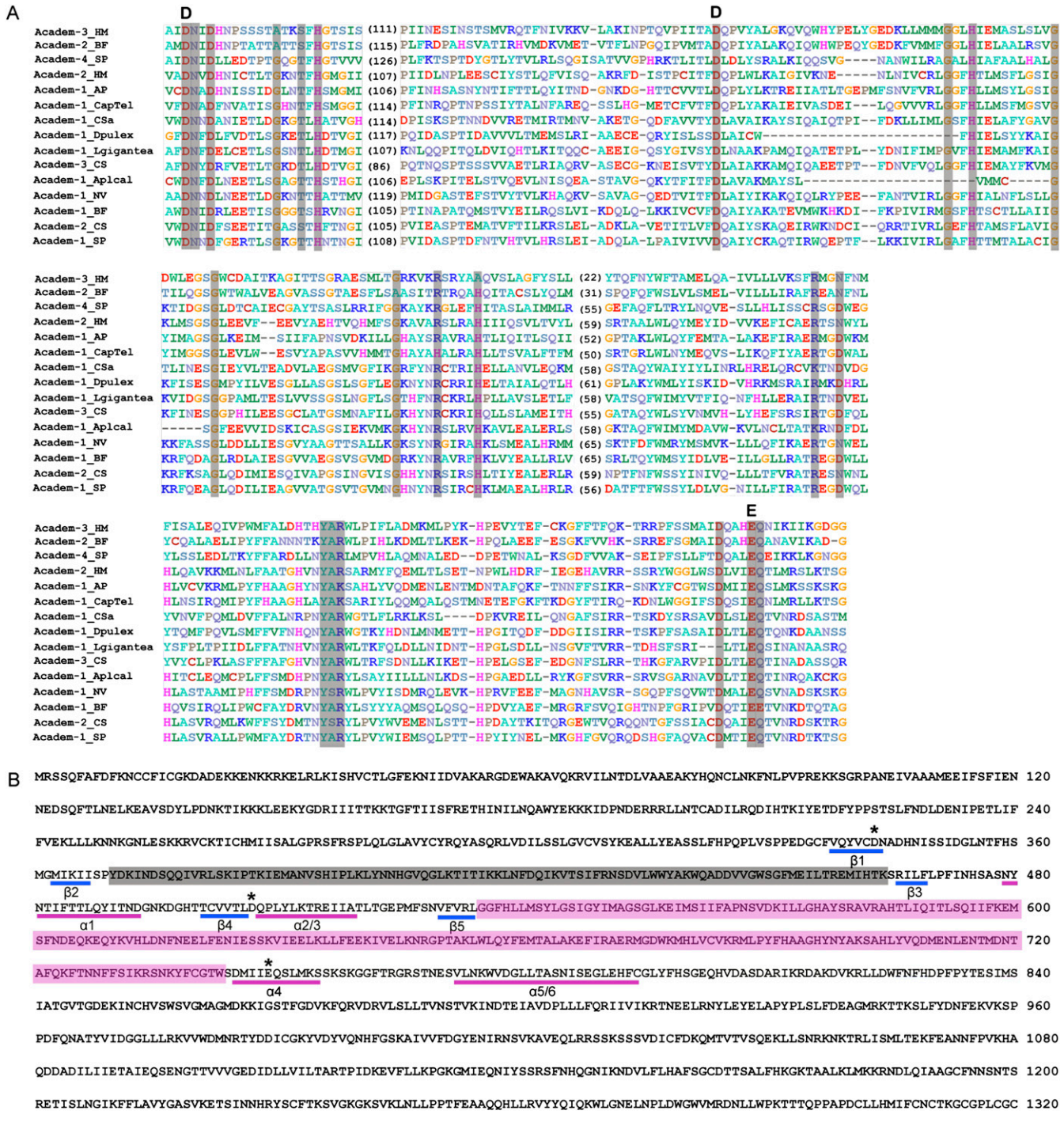

Fig. S5. The DDE domain of *Academ* elements. (*A*) Alignment shown is after redundancy elimination. Conserved residues are highlighted in gray. The DDE triad is marked above the alignment. Distances between the conserved blocks are indicated in number of amino acid residues. (*B*) Predicted secondary structure of the DDE domain of the *Academ-1_AP* transposase. The DDE triad is indicated by asterisks. α-Helices and β-strands are highlighted with pink and blue bars, respectively. The first D is on β1, the second D is right after β4, and the E is on α4. The inserted domain between β5 and α4 (highlighted in pink background) is entirely α-helical. An additional inserted domain between β2 and β3 is predominantly coiled (highlighted in gray).

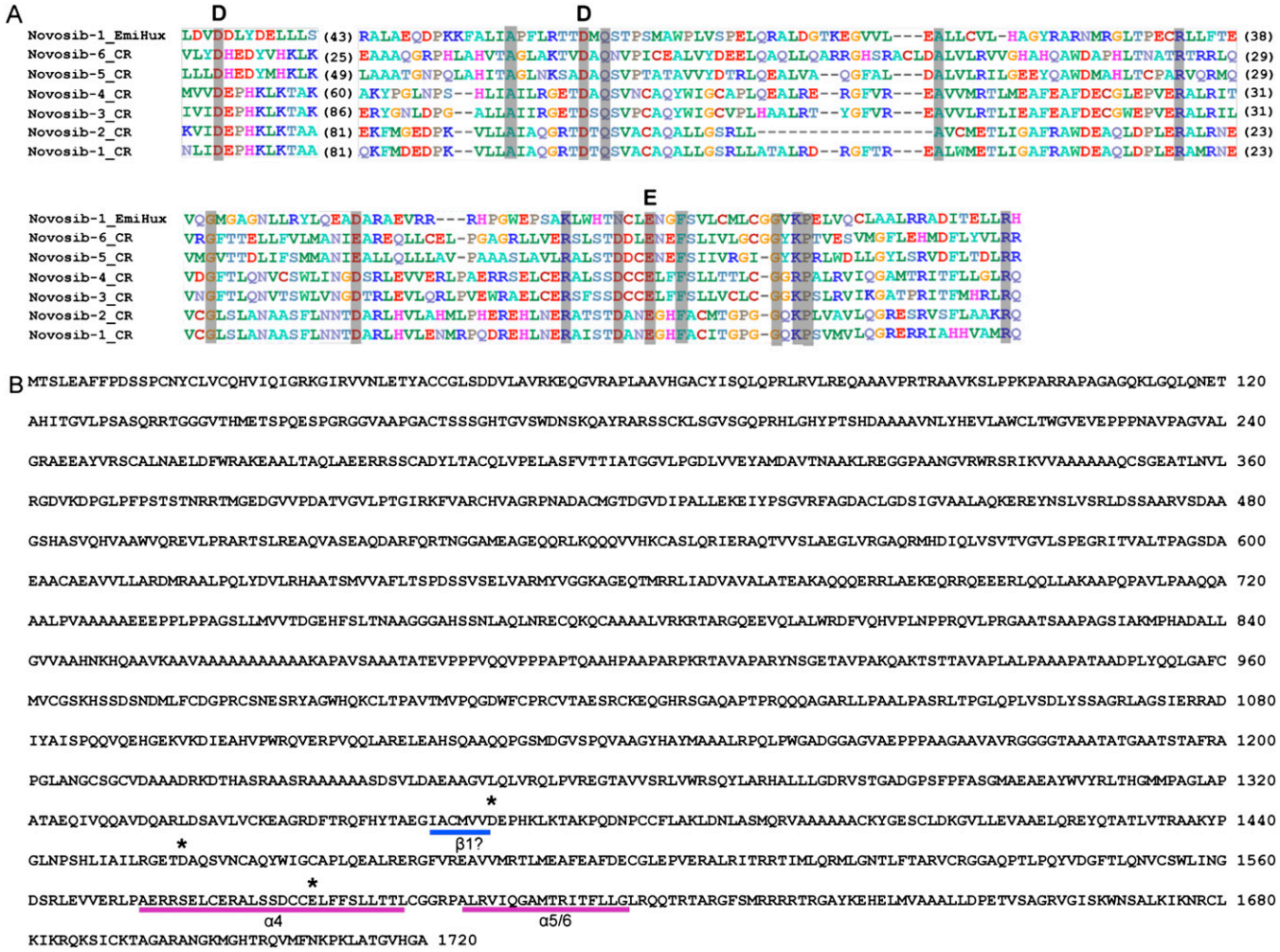

Fig. S6. The putative DDE domain of *Novosib* elements. (*A*) Alignment shown is before redundancy elimination, because there are only seven elements available. Conserved residues are highlighted in gray. The tentative DDE triad is marked above the alignment. Distances between the conserved blocks are indicated in number of amino acid residues. (*B*) Predicted secondary structure around the tentative DDE triad (marked by asterisks) of the *Novosib-1_CR* transposase. The typical "β1-β2-β3-α1-β4-α2/3-β5-α4-α5/6" fold of the DDE domain is not yet confirmed for the available *Novosib* transposases.

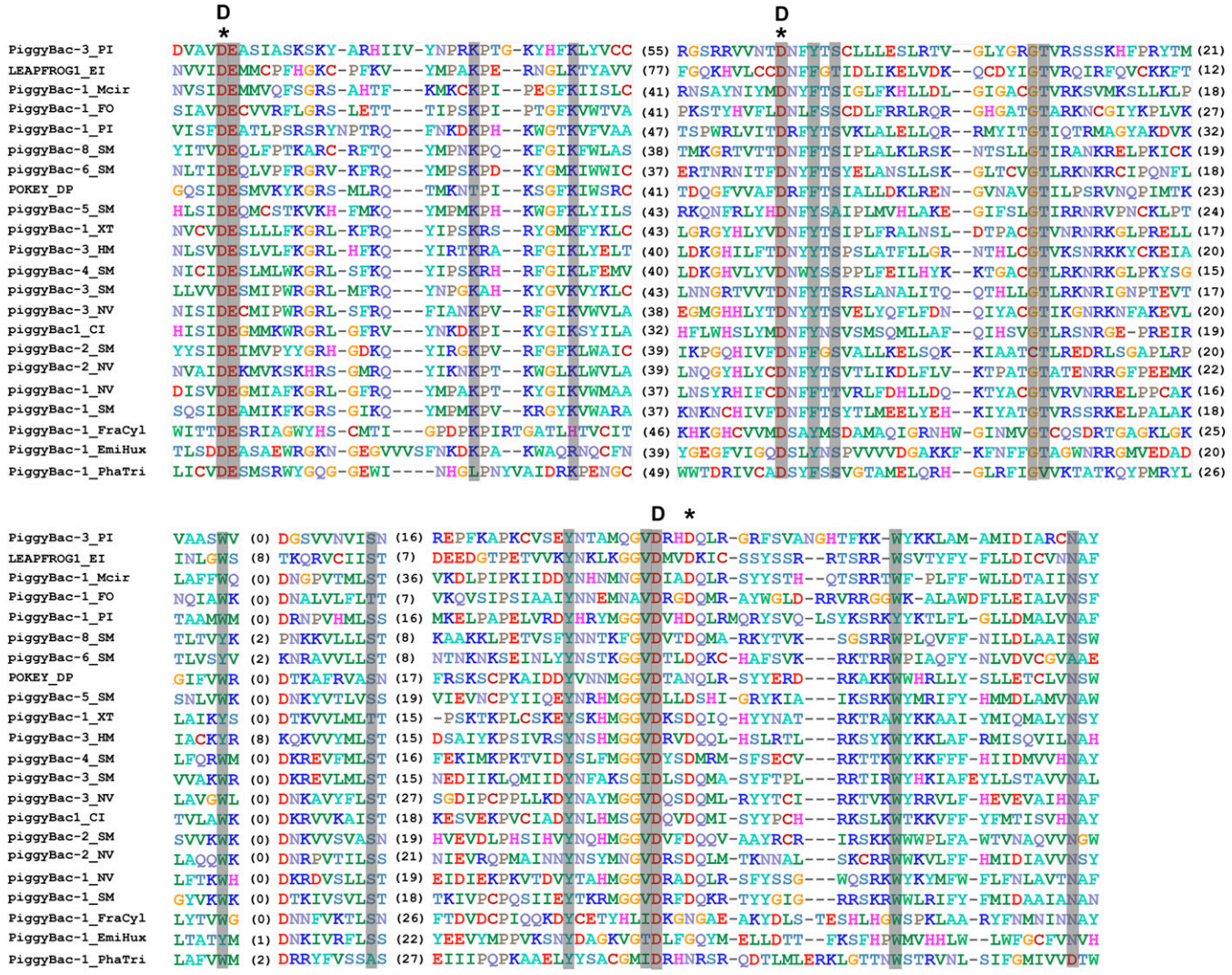

Fig. S7. The DDD domain of *PiggyBac* elements. Alignment shown is after redundancy elimination. Conserved residues are highlighted in gray. The DDD triad identified here and by a previous experimental study (1), corresponding to D268, D346, and D447 of the *Trichoplusia ni PiggyBac* transposase (GenBank: AAA87375.2), are marked with letters above the alignment. The DDD triad suggested in Bao et al. (2) are marked with asterisks. Distances between the conserved blocks are indicated in number of amino acid residues.

1. Mitra R, Fain-Thornton J, Craig NL (2008) *piggyBac* can bypass DNA synthesis during cut and paste transposition. *EMBO J* 27:1097–1109.
2. Bao WD, Jurka MG, Kapitonov VV, Jurka J (2009) New superfamilies of eukaryotic DNA transposons and their internal divisions. *Mol Biol Evol* 26:983–993.

Table S1. Summary of the TSD feature, terminal motif of the terminal inverted repeat, and the signature string of each superfamily in the revised system

Superfamily	TSD	Terminal motif	Signature string
Tc1/mariner	TA	Variable	**D**[E/q](80-125)**D**[N/G/s](3)H(15-26)P(9-10)[**D/E**]
Zator	TWA	5′-GGS	**D**[D/q]K(~30)DH(8-11)[P/s](~60)**D**[G/n]G(~17)*F*(5)D(8)[P/a](2)S(2)N(2)**E**
Merlin	8–(9) bp	5′-GG	**D**E(6-7)[K/r](3)[G/a](5-8)W(2)[G/t](~16)R(8)[I/l](7)[T/S](1)[I/v](1)[T/S] **D**(4)Y(~10)H(4)H(3)[F/y](~11)[I/v]**E**(2)W(3-4)K
PIF/Harbinger	TWA, AT or AWT	Variable	**D**(8-9)P(~40)[G/a](3)D(21-42)**D**(1)[G/A/i][F/Y](~30)R(2)[V/I]**E**(7-8)[F/W/y]
MULE	9–10 bp	Variable	**D**(1)[T/a/c](~60)**D**(7)[A/g/s](10-28)[C/d](2)H(50-160)[W/f](~16)[N/S/*T*](3)**E**(2)[F/N/H]
P	7–8 bp	5′-CANRG	**D**(~90)**D**(4)[N/y/h](~30)D(2)H(2)K(2)[R/y][N/t](~190)**E**(2)[F/n](3)R
hAT	(5–7)–8 bp	5′-YARNG	**D**(50-90)**D**(15-45)C(2)H(40-120)R[W/f](~300)**E**(2)[F/w/V/i]
Kolobok	TTAA	5′-RG	**D**(45-75)[M/I]E(~25)**D**(18-24)[C/D](2)H(~65)H(7-24)C(~60)**E**(2)[N/H/f](8-10)K
Novosib	8 bp	Variable	**D**(50-110)A(6)**D**(1)Q(~40)A(10)R(~35)G(13)[D/E](~19)[R/k](4)[D/n](2)**E**(2)*F*(7-8)G(1)[K/r]P(18)R
Sola1	WWWW	Variable	**D**(6-7)P(~70)**D**(4)QN[K/R][N/c/s](~22)[F/Y](3)GH(5)**D**
Sola2	4 bp	5′-GRG	**D**F[A/S]E(10)Q(3)[F/W/y](~27)[S/t](~32)[S/t]**D**G(3)Q[Y/F][K/R][N/c/s](~19)W(1)[F/y](4)HGK(3)**D**(6)K
Sola3	TTAA	5′-SRG	**D**[F/W/y](2)K(6)RE(5)[F/y]GK(1)G(3)H(~60)**D**[N/g](3)Y(~30)[G/n]K(3)**D**(3)A
PiggyBac	TTAA	5′-CMY	**D**E(~17)[K/t/l](6)[K/r/h](~50)**D**(2)[F/Y](1)[S/t/a](15)GT(~35)[W/Y](~10) [S/t/a](~30)Y(6)[V/i]**D**(~17)[W](~14)[N/d/a]
CMC	2–4 bp	5′-CMC	**D**(70-100)**D**(17-21)C(2)C(50-110)[M/L]H(3-4)G(140-225)H(3-4)H(19-20)**E**
Transib	5 bp	5′-CACWNTG	G(1)**D**G(~55)CRP(7)E(~42)**D**(1)K(~13)C(2)C(~30)LH(~60)**GN**(2)[R/k](~65)H(2)[L/i](1)H(20)**E**(3)K(5)R
Academ	3 bp	5′-TAG	**D**N(1)D(7)[G/A](2)[T/S](1)H(~145)**D**(~25)G(2)H(9)G(6)G(~21)[G/a](4)[R/h](4)H(~95) [R/k](2)[D/N](22)Y[A/s][R/k](~42)D(3)**E**[Q/e]
Ginger	4–(5) bp	5′-TGT	**D**(~20)[S/T][K/r][W/F/Y](8)[K/r](~16)G(7)**D**(1)G(1)EF(~21)[P/s](2)[Q/N]G(2)**E**(2)N

Parentheses in the TSDs indicate exceptional cases. The DDE/D triads are boldface. Alternative residues are marked by slashes; lowercase indicates that a residue occurs in <10% of the sequences in the alignment profile.

Table S2. Taxonomic distribution of the 17 superfamilies across five eukaryotic supergroups

		Opisthokonts				
Superfamily	Plantae	Animals	Fungi	Amoebozoa	Excavates	Chromalveolates
Tc1/mariner	Y	Y	Y	Y	Y	Y
Zator		Y			Y	
Merlin	**Y**	Y	**Y**			Y
PIF/Harbinger	Y	Y	Y	**Y**	**Y**	Y
MULE	Y	Y	Y	Y	Y	Y
P	Y	Y	**Y**		Y	**Y**
hAT	Y	Y	Y	Y	Y	Y
Kolobok	**Y**	Y			**Y**	**Y**
Novisib	Y					**Y**
PiggyBac	**Y**	Y	Y	Y	Y	Y
Sola1	Y	Y			Y	Y
Sola2	Y					Y
Sola3	Y					Y
CMC	Y	Y	Y		**Y**	
Transib		Y				
Academ		Y				
Ginger	**Y**	Y		Y		Y

The supergroup Opisthokonts is divided into animals and fungi. "Y" indicates presence and blank indicates absence; boldface indicates presence unknown before this study.

Other Supporting Information Files

Dataset S1 (TXT)