Supporting information.

Supp. Table 1: Distribution of fungal TET/JBP, KDZP transposases and KDZP transposons in eukaryotes

Species name	Species abbrevi-ations	Number of DNMT1 (D), RID1 (Ri), RAD5-fused-DCM (Ra), novel adenine methylase (Am)	Active/Total [a] (TET/JBP)	Number of Kyakuja (K) Dileera(D), Zisupton (Z), Plavaka (P)	Number of TET/JBP-associated with K,D,Z or P	Other transposon or TET/JBP- associated domains
Agaricomycetes						
Auricularia delicata TFB-10046 SS5	Adel	2D, 1Ra	60/82	15K, 86D, 110Z, 69P	7K, 21Z	AdelX1, AdelTA2, SMT4, SWIM
Laccaria bicolor S238N-H82	Lbic	2D, 1Ra	47/74	54K, 39D, 51Z, 79P	15K	HMG, SMT4, SWIM
Coprinopsis cinerea okayama7#130	Ccin	2D, 1Ra	32/47	50K, 14D, 6Z, 30P	37K	HMG, cys clus, alpha-helical domain, AHD2, AHD3, SWIM
Stereum hirsutum FP-91666 SS1	Shir	2D, 1Ra	27/39	12K, 4D, 5Z, 27P	6K, 1D	SMT4
Fomitopsis pinicola	Fpin	2D, 1Ra	19/41	30K, 18D, 2Z, 85P	3K, 12P	
Serpula lacrymans var. lacrymans S7.3	Slac	2D, 1Ra	11/38	10K, 27D, 5Z, 168P	1K, 5P	SMT4
Serpula lacrymans var. lacrymans S7.9	Slac	1D	7/34	7K, 13D, 4Z, 95P	3P	SMT4
Coniophora puteana RWD-64-598 SS2	Cput	3D, 1Ra	6/29	41K, 38D, 20Z, 115P	9K	HMG
Agaricus bisporus var. burnettii JB137-S8	Abis	1D, 1Ra	14/18	2K, 1D, 5Z, 22P	2Z, 1P	SMT4, SWIM
Agaricus bisporus var. bisporus H97	Abis	2D, 1Ra	10/18	4K, 2D, 8Z, 42P	4P	SWIM
Schizophyllum commune H4-8	Scom	2D, 1Ra	12/16	18K, 7D, 1Z, 26P	5K, 3D	
Ceriporiopsis subvermispora B	Csub	2D, 1Ra	9/12	27K, 1Z, 44P	4K	SMT4
Trametes versicolor FP-101664 SS1	Tver	2D, 1Ra	7/11	21K, 10D, 10Z, 34P	4K, 2P	SMT4
Dichomitus squalens LYAD-421 SS1	Dsqu	2D, 1Ra	5/6	7K, 1D, 3Z, 35P		

Punctularia strigosozonata HHB-11173 SS5	Pstr	1D, 1Ra	9/16	25K, 23D, 5Z, 13P	1Z	
Moniliophthora perniciosa FA553	Mper	2D, 1Ra	4/11	34K, 1Z, 7P		
Rhizoctonia solani AG-1 IA	Rsol	1D, 1Ra	?	1P		
Rhizoctonia solani AG-1 IB	Rsol	1D, 1Ra	1/1	2K, 1D, 9P		
Fomitiporia mediterranea MF3/22	Fmed	3D, 1Ra	?	1K, 2D, 17P		
Phanerochaete carnosa HHB-10118-sp	Pcar	3D, 1Ra	1?/2	10K, 8D, 6Z, 20P		SMT4
Postia placenta Mad-698-R	Ppla	2D, 1Ra	0/1	1K, 11D, 12P		SMT4
Piriformospora indica DSM 11827	Pind	2D, 1Ra	1/1	6K, 6P		
Dacryomycetes						
Dacryopinax sp. DJM-731 SS1	Dsp	2D, 1Ra	¾	5K, 22D, 11Z, 13P		SMT4
Tremellomycetes						
Tremella mesenterica DSM 1558	Tmes	1Ra	15/23			
Tremella fuciformis	Tfuc	?	1/1			
Trichosporon asahii var. asahii CBS 8904	Tasa	1D, 1Ra	4/4			
Trichosporon asahii var. asahii CBS 2479	Tasa	1D, 1Ra	2/2			
Cryptococcus neoformans var. grubii H99 *	Cneo	1Ra	1/1	3D, 1Z, 1P		
Cryptococcus gattii WM276	Cgat	1Ra	?	1D		
Cryptococcus neoformans var. neoformans B-3501A	Cneo	1Ra	?			
Pucciniomycotina -> Pucciniomycetes						
Puccinia graminis f. sp. tritici CRL 75-36-700-3	Pgra	1D, 1Ra	103/215	2K, 91D, 1Z, 5P	14D, 1P	HMG
Melampsora larici-populina 98AG31	Mlar	1D, 1Ra	31/115	73D, 29Z,	14D, 1Z	HMG, SMT4
Pucciniomycotina -> Microbotryomycetes						
Rhodotorula glutinis ATCC 204091	Rglu	1D, 1Ra	3/5			
Rhodosporidium toruloides NP11	Rtor	1D, 1Ra	4/4			
Pucciniomycotina -> Mixiomycetes						
Mixia osmundae IAM 14324	Mosm			2Z		SWIM
Ustilagomycotina						

Sporisorium reilianum SRZ2	Srei	1D, 1Ra	4/6			
Ustilago hordei	Uhor	1D, 1Ra	2/6			
Saccharomycotina-> Pezizomycotina						
Eutypa lata UCREL1	Elat	1D, 1Ra, 1Ri	12/15			
Geomyces destructans (also *Pseudogymnoascus destructans*) 20631-21	Gdes	1D, 1Ra, 1Ri	2/5			
Zymoseptoria tritici IPO323	Ztri	1Ra	1/2			
Mycosphaerella populorum SO2202/ Sphaerulina musiva SO2202	Mpop	1D, 1Ra, 1Ri	0/1			
Glarea lozoyensis 74030	Gloz	1D, 1Ri	1/1			
Coccidioides immitis RS	Cimm	2D, 3Ra, 2Ri		**1P**		
Togninia minima UCRPA7	Tmin		0/1			
Taphrinomycotina						
Schizosaccharomyces pombe 972h-	Spom		0/1			
Schizosaccharomyces japonicus yFS275	Sjap		0/1			
Glomeromycota						
Rhizophagus irregularis (formerly Glomus intraradices)	Rirr	4D, 1Am[b]	4/5	**5D, 18P**		
Mucormycotina						
Mortierella verticillata	Mver	1D, 1Am[b]	1/1			
Phycomyces blakesleeanus	Pbla	2D, 1Am[b]		**9D**		
Mucor circinelloides f. circinelloides	Mcir	1Am[b]		**6D**		
Rhizopus delemar	Rdel	1Am[b]		**25P**		
Chytridiomycota->Blastocladiales						
Allomyces macrogynus	Amac	4D		**2Z**		
Chytridiomycota->euchytrids						
Spizellomyces punctatus	Spun	1Ra, 1Am[b]				
Viridiplantae -> Chlorophyta						
Chlamydomonas reinhardtii	Crei	1D[c]	8/9	**5K**, **1D**, **5Z**		

Volvox carteri f. nagariensis	Vcar	1D[c]	½	**7K**, **1Z**		
Chlorella variabilis	Cvar	1D[c]		**2K**		
Ostreococcus tauri	Otau	1R[c]		**1K**		
Viridiplantae -> Streptophyta						
Physcomitrella patens subsp. patens	Ppat	1D[c]	0/3			
Amoebozoa						
Acanthamoeba castellanii str. Neff	Acas	5D	1/2			
Metazoa						
Hydra magnipapillata	Hmag	1D[c]	1/1 [d]	**1D**, **1Z**		
Nematostella vectensis	Nvec	1D[c]	1/1 [d]	**3K**		
Amphimedon queenslandica	Aque	1D[c]	1/1 [d]	**3K**,**2D**, **2Z**		
Capitella teleta	Ctel	1D[c]	1/1 [d]	**5K**, **3D**, **1Z**		
Crassostrea gigas	Cgig	1D[c]	1/1 [d]	**1K**, **1D**		
Branchiostoma floridae	Bflo	1D[c]	1/1 [d]	**1K**, **6Z**		
Daphnia pulex	Dpul	1D[c]	1/1 [d]	**27K**		
Saccoglossus kowalevskii	Skow	1D[c]	1/1 [d]	**1Z**		SWIM, SMT4, F-box, SAP
Danio rerio	Drer	1D[c]	3/3 [d]	**2K, 7Z, 1Z*** [e]		SWIM, SMT4, F-box, SAP
Oreochromis niloticus	Onil	1D[c]	3/3 [d]	**1K, 1Z, 1Z*** [e]		SWIM, SMT4, F-box, SAP
Oryzias latipes	Olat	1D[c]	3/3 [d]	**1K, 1Z*** [e]		SWIM, SMT4, F-box, SAP, HMG
Xiphophorus maculatus	Xmac	1D?[c]	3/3 [d] ?	**1Z**		SWIM, SMT4, F-box, SAP
Tetraodon nigroviridis	Tnig	1D[c]	3/3 [d]	**1Z*** [e]		HMG
Homo sapiens and other tetrapods	Hsap	1D[c]	3/3 [d]	**1Z*** [e]		HMG
Stramenopiles						
Thalassiosira oceanica	Toce	1D[c], 1Ra	2 [d]	**2P**		
Haptophyceae						

Emiliania huxleyi CCMP1516	Ehux	1D[c], 2Ra	5[d]	1Z		

a: When the chromosomes are not assembled completely, the total numbers here are reported based on the sequenced contigs. These numbers could hence be the lower bound rather than the exact number in the complete genome.

b: Except for *Rhizopus irregularis*, all novel adenine methylase containing species have the following architecture: chromo(2-3)+ZZ+PHD+PHD+ZZ+GATA+DAM. The *Rhizophagus irregularis* architecture is GATA+DAM. For a list of gis, please refer to the html page at the ftp site listed below.

c: These species have several other families of DNA methyltransferases, in addition to DNMT1

d: These families are distinct from the fungal type TET/JBP family

e: These are the HMGXB3-like domesticated Zisupoton transposases.

For a complete list of gis and gene neighborhoods of TET/JBP and KDZP transposases, please refer to ftp://ftp.ncbi.nlm.nih.gov/pub/aravind/temp/fungaltettransposon/fungaltets.html

Note: There is a nearly absolute correlation between the presence of TET/JBP and DNMT1/RID1 genes in the fungi. The only exceptions in the currently available data are seen in *Tremella mesenterica*, *Tremella fuciformis* and *Cryptococcus neoformans* var. *grubii* H99 in basidiomycetes and *Zymoseptoria tritici* in ascomycetes, which contain active TET/JBP genes but no DNMT1. While these exceptions are intriguing, it should be noted that in each of these cases the genome assembly available in Genbank is not complete with several contigs whose mapping to chromosomes being as yet undetermined. Hence, we refrain from any evolutionary or functional inferences from these observations. However, three of the above species (except Tremella fuciformis), contain a Rad5-fused DNA cytosine methylase, which in principle could generate methyl cytosine substrates for further oxidation in these species.

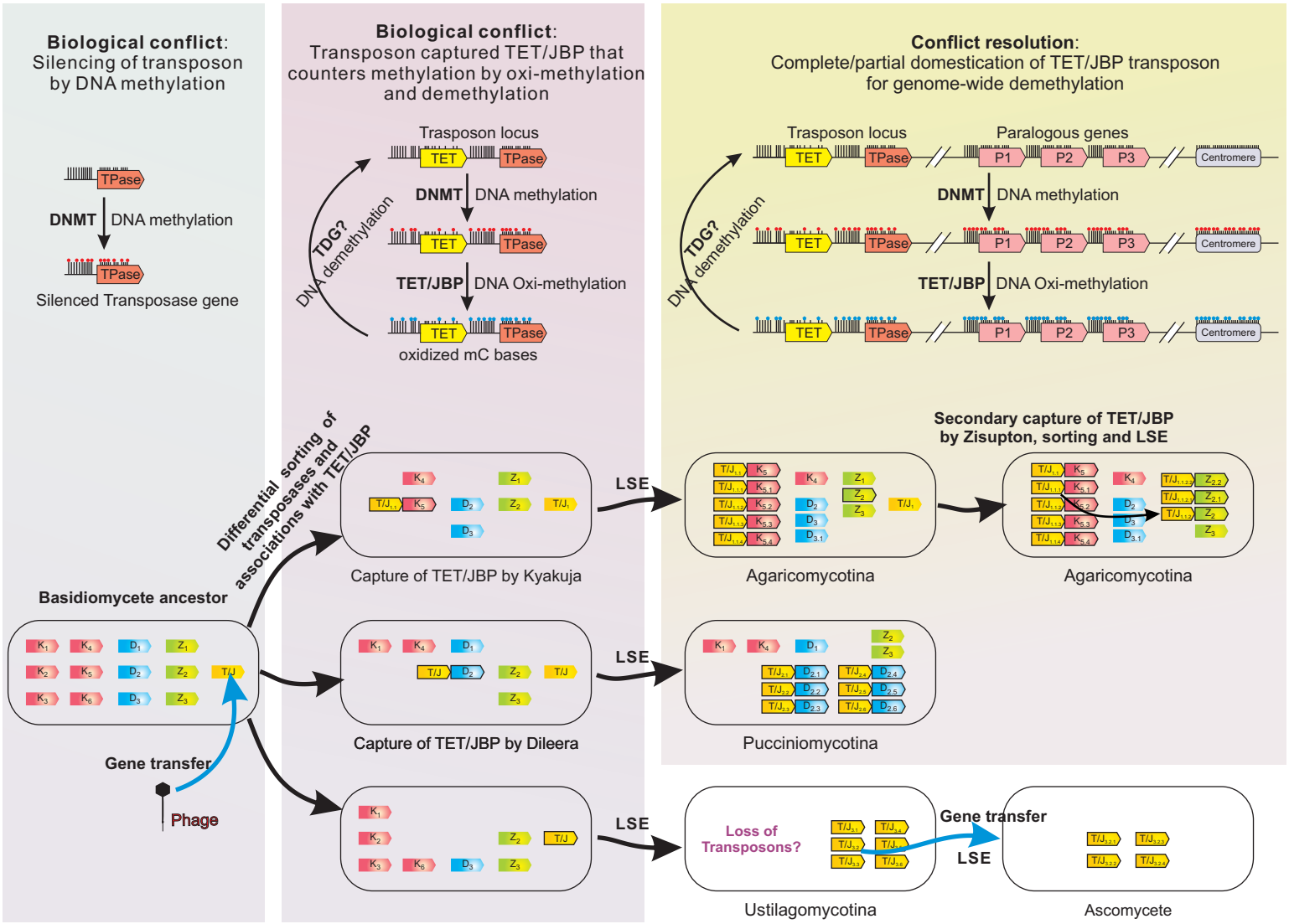

Fig. S1. Model for transposon inheritance and nature of biological conflict. The model illustrates the concept of assortative inheritance of KDZ transposons, shaped by loss, lineage-specific expansions (LSE) and associations with TET/JBP genes. Also shown is the nature of the biological conflict in the presence and absence of TET/JBP genes.

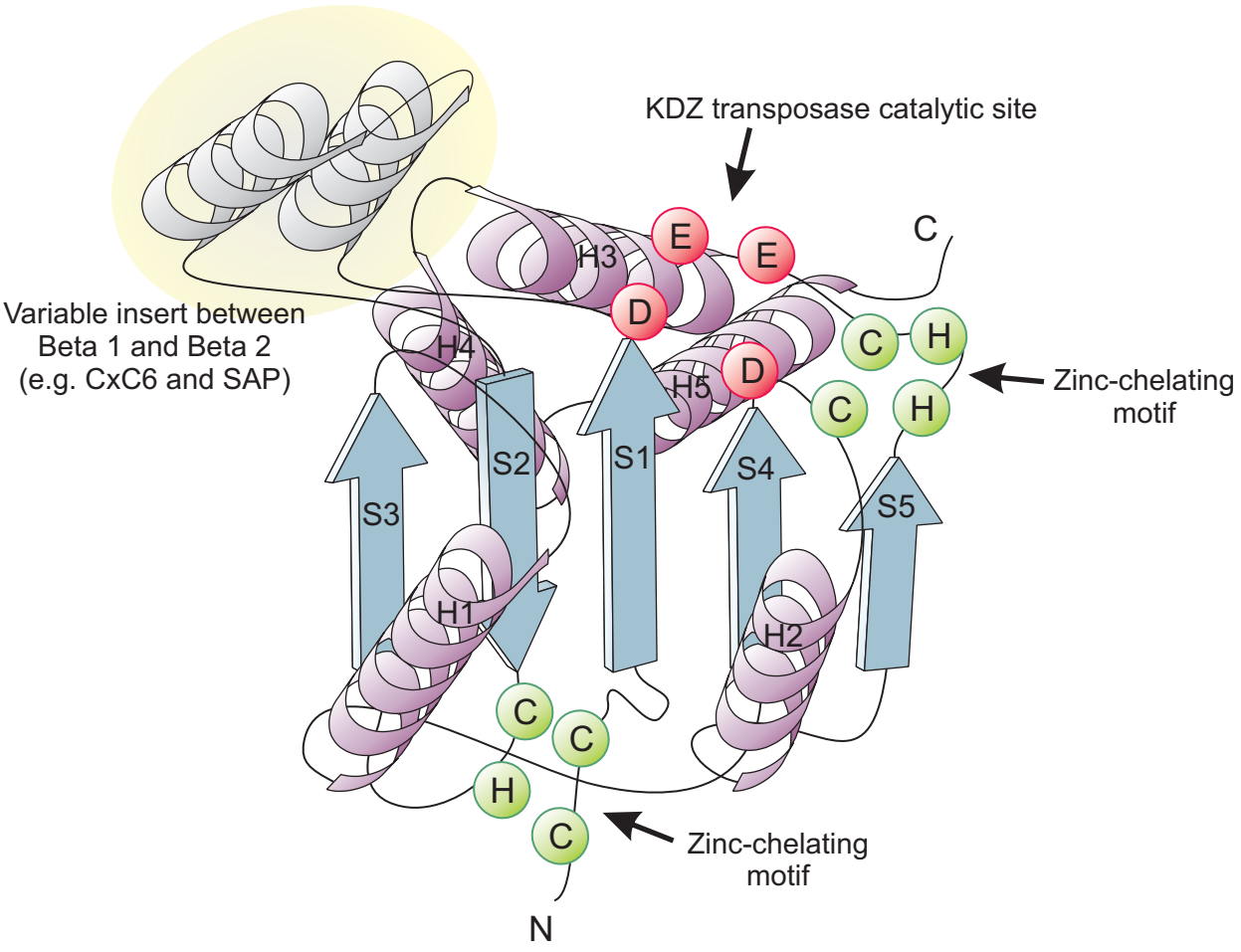

Fig. S2. Topology of the KDZ transposase domain. Conserved five beta-strands (S1-5) and five alpha-helices (H1-5) are shown in grey blue and purple, respectively. Catalytic site residues are shown in red balls while two Zinc-chelating motifs are shown in green balls. Variable insert between S1 and S2 is highlighted in yellow circle.

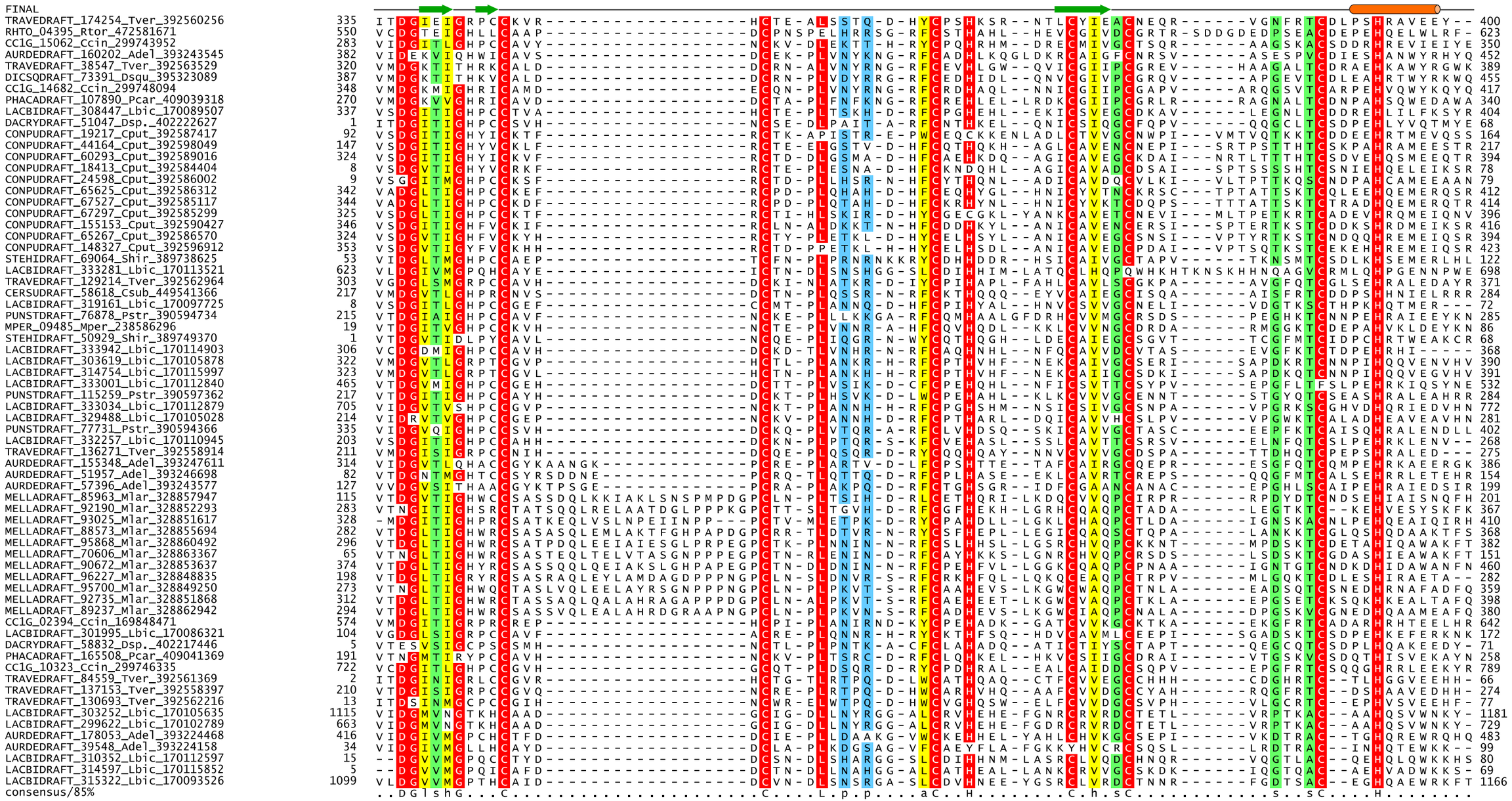

Fig. S3. Multiple sequence alignment of the CxC6 domain.

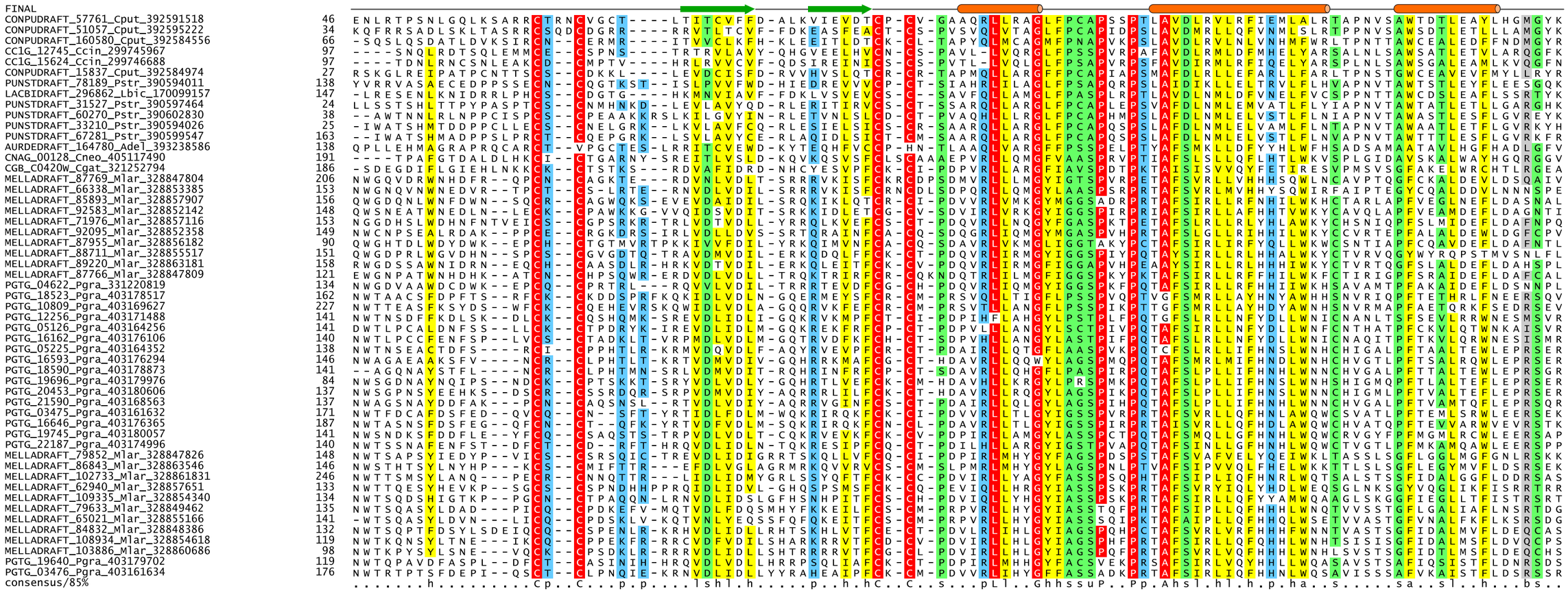

Fig. S4. Multiple sequence alignment of the CxC1 domain.

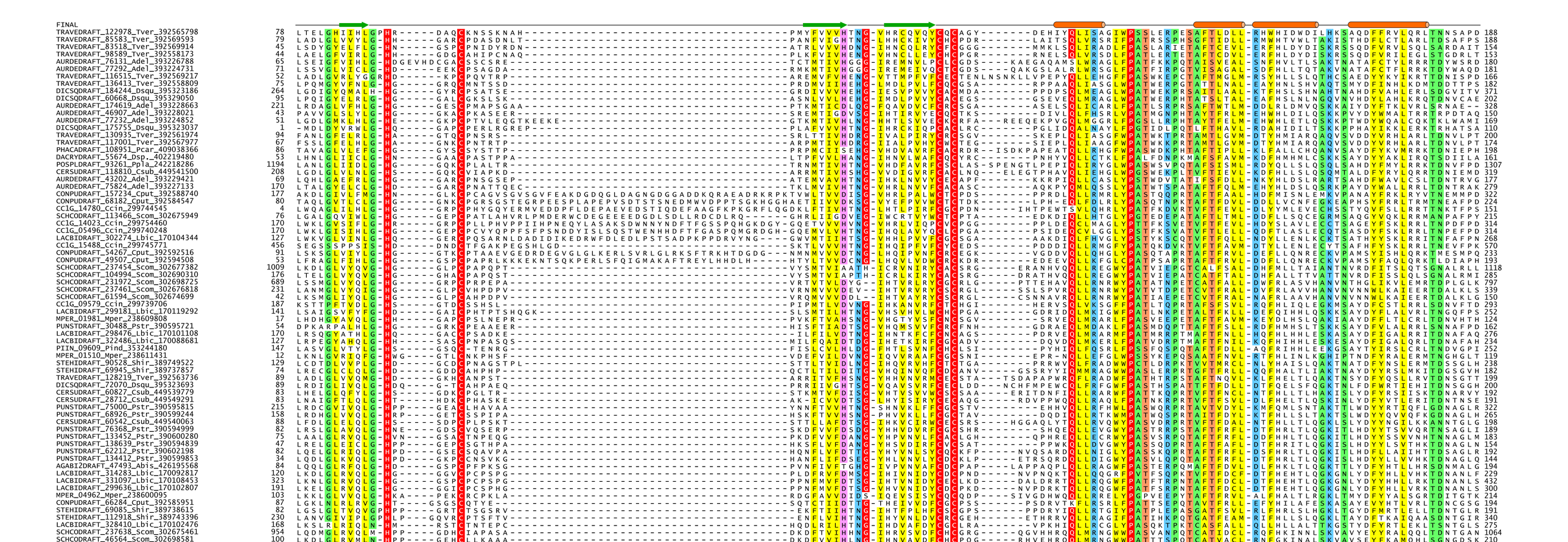

Fig. S5. Multiple sequence alignment of the CxC2 domain.

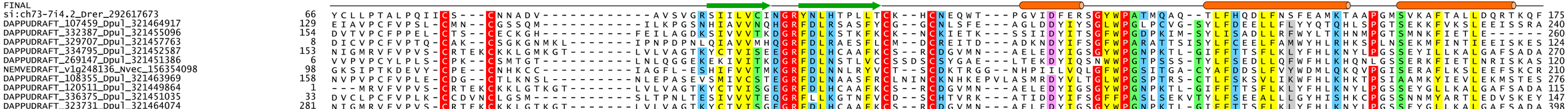

Fig. S6. Multiple sequence alignment of the CxC3 domain.

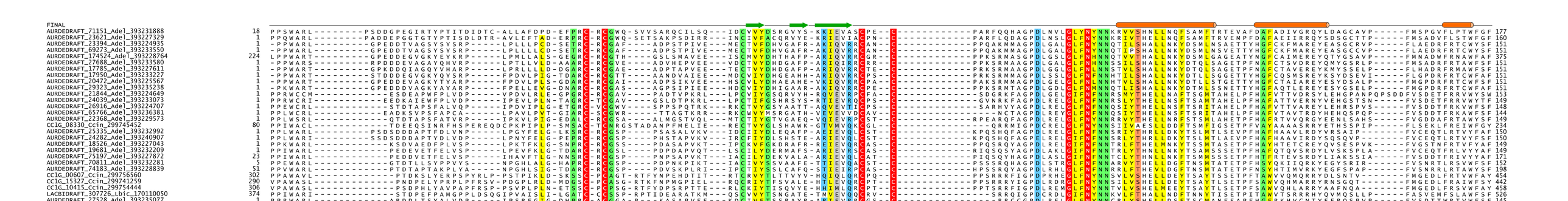

Fig. S7. Multiple sequence alignment of the CxC4 domain.

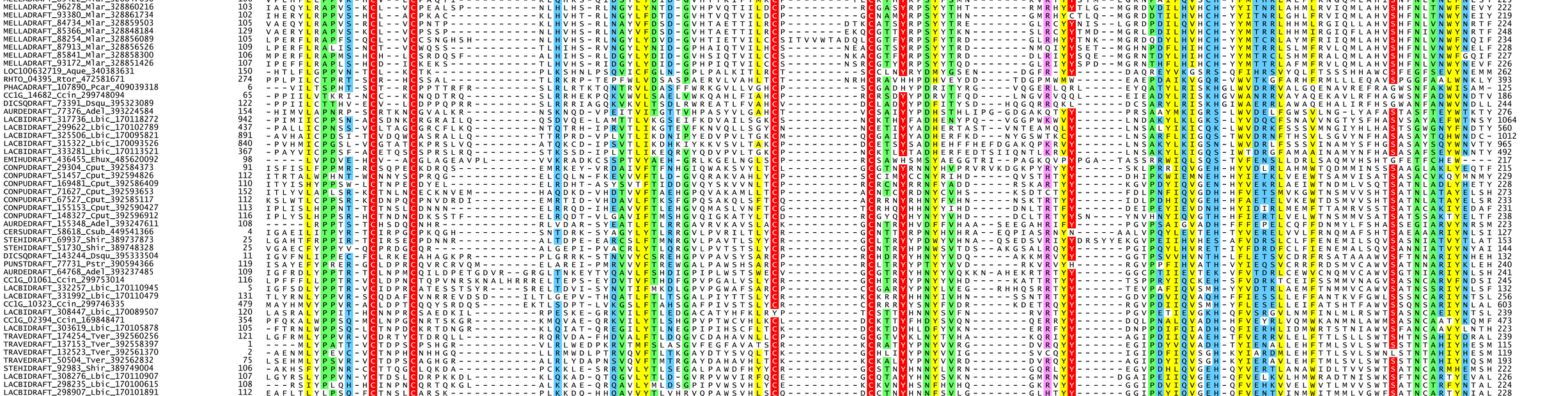

Fig. S8. Multiple sequence alignment of the CxC5 domain.

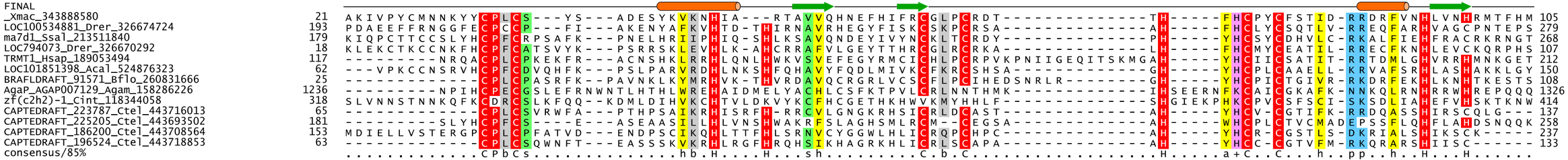

Fig. S9. Multiple sequence alignment of the TRMT1N domain.

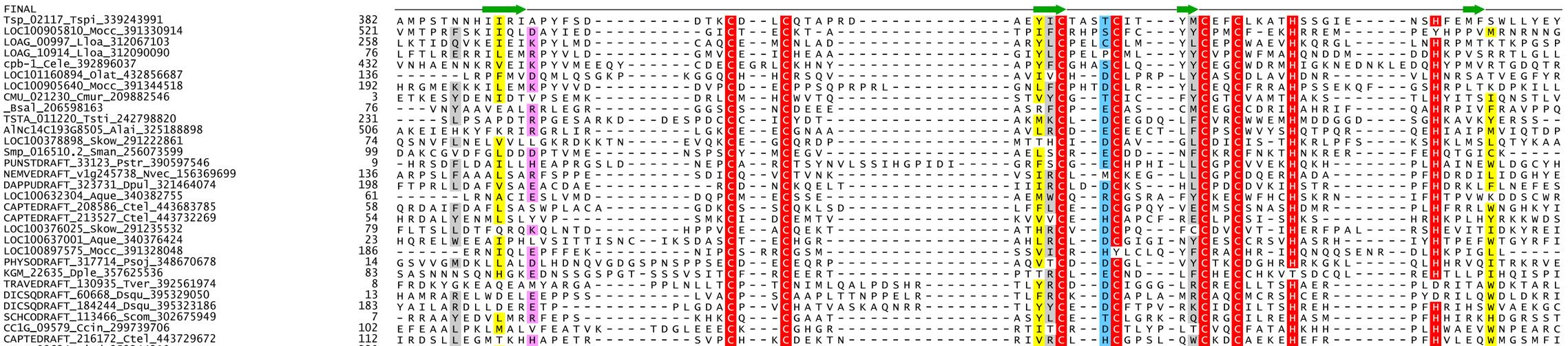

Fig. S10. Multiple sequence alignment of the ZZ-finger domain.

```
FINAL
_Xmac_343888580                    1333  NWENFSESFKAENRIPFAVIQCSSCNRQYKD-SVGFTGHGRRGEFPRYYCD-GIPGHC-PECQY 1393
LOC567404_Drer_292613057            105  NWSRFSEDYKKEFRVPYSLTECPECGDIFKDCPPGYVGDGRKGVLRGFYSTIDFPGYCSAECHF 168
LOC100332256_Drer_292612984         105  NWSRFSEDYKKEFRVPYSLTECPECGDIFKDCPPGYVGDGRKGVLRGFYSTIDFPGYCSAECHF 168
si:ch211-238e22.4_Drer_292613145    116  NWSRFSEDYKKEFRVPYSLTECPECGDIFKDCPPGYVGDGRKGVLRGFYSTIDFPGYCSAECHF 179
LOC100333706_Drer_292613241         105  DWSRFSEDYKKEFRVPYSLTECPECGDIFKDCPPGYVGDGRKGVLRGFYSTIDFPGYCSAECHF 168
si:dkey-122c11.1_Drer_190194295     105  NWSRFSEDYKKEFRVPCSLTECPKCGDIFKDCPPGYVGDGRKGVLRGFYSTIDFPGYCLAECHF 168
LOC100692678_Onil_348544903         571  -WSSFSEKYKEQYFNMYTIEICLQCGDQYKHCPRGYVGTGKRGQLRAFYSEEMRPGYCS----- 628
consensus/85%                            sWSpFSEcYKcEaR]Pao]pbCspCGDbaKDCPsGYVGcGR+G.LRuFYSsbshPGYC.sECpa
```

Fig. S11. Multiple sequence alignment of the CxC7 domain

```
FINAL
AURDEDRAFT_177778_AdeI_393224953   210  FYPLVRK--AYAMNRFRESYVAVIAMPVSGDLKGQTYAWGSPRLLQDLPDFLDDCAFEVSHGTLDKRAHGYAKA 281
AURDEDRAFT_176894_AdeI_393226259   437  FYPLVRK--GYAMNRFRESYVAIIAMPVSGDLKGQTYAWGSPRLLQDLPDFLDDCAFEVSHGTMDKRAHGYARA 508
AURDEDRAFT_175055_AdeI_393228199    34  --PLVRKLKDYAMNRFRESYVAVIAMPVSRDLKGQTYAWGSPRLLQDLPEFLDDCAFEVSHGTMDKRVHGYAKA 105
AURDEDRAFT_178703_AdeI_393223315   341  FYPLVRK--AYAMNRFRESYVAVIAMPVSGDLKGQTYAWASPRLLQDLPDFLDDCALEISHGTMDKRAHGYAKA 412
AURDEDRAFT_175659_AdeI_393227612   427  FYPLVRK--AYAMNRFRESYVAVIAMPVSGDLKGMTYAWGSPRLLQDLPDFLDDCAFEIAHGTMDKRAHGYTKA 498
AURDEDRAFT_176752_AdeI_393226440   450  FYPLVRK--AYAMNRFRESYVAVIAMPVSGDLKGQTYAWGSPRLLKDLPDFLDDCAFEIAHGTMDKRAHGYAKA 521
AURDEDRAFT_176345_AdeI_393226915   398  FYPIVRK--AYAMNRFRESYVAVIAMPVSGDLKGQTYAWGSPRLLKDLPDFLDDCAFEISHGTMDKRAHGYAKA 469
AURDEDRAFT_173009_AdeI_393230259   423  FYPIVRK--AYAMNRFRESYVAVIAMPVSGDLKGMTYAWASPRLLQDLPDFLDDCVFEISHGTMDKRAHGYAKA 494
AURDEDRAFT_169741_AdeI_393233581   430  FYPLVRK--GYAMNRFRESYVAIIAMPVSGDLKGQTYAWASPRLLQDLPDFLDDCALELSHGTLDKRAHGYAKA 501
AURDEDRAFT_175063_AdeI_393228207   424  FYPIVRK--AYAMNRFRKSYVAVIAMPVSGDLKGMTYAWASPRLLQDLPDFLDDCAFEISHGTMDKRAHGYAKA 495
AURDEDRAFT_177075_AdeI_393226020   419  FYPLVRK--AYAMNRYRESYVAVIAMPVSGDLKGMTYAWGSPRLLKDLPDFLDDCAFEVSHGTMDKRAHGYAKA 490
AURDEDRAFT_166204_AdeI_393237175   408  FYPLVRK--GYAMNRYRESYIAIIVMPVSGDLKGQTYAWGSPRLLQDLPEFLDDCALEISHGTMDKRAHGYAKA 479
AURDEDRAFT_131245_AdeI_393226547   715  FYAAARK--GYAMNNFLESYVLIMAMPTLGDMKGQTYTWASPRLIQDAPHIIDQAAYDLAKDTMDKRAHGYAKA 786
AURDEDRAFT_170002_AdeI_393233452   523  FYPIARK--SFAMNKSRESYILTLAMPLGSDVGGQTYAWASPRLMEDMPELLDEVVWRVANHTADKRANGYKQ- 593
AURDEDRAFT_170219_AdeI_393233228   516  FYPIARK--GYAMNTFQPSYVLVVAMPIKPELKSQIYTWASDSLIQDAPQLIGEAAHQLAHRTAHKRAHGAAR- 586
AURDEDRAFT_177385_AdeI_393225566   543  FYAVARK--AFAMNNFCESYIVVMALPLAGDLRGQTYVWSSPRVEQDLPGFVDDAAHELNVSSWVHRQH----- 609
AURDEDRAFT_127329_AdeI_393233551   603  FYPIARK--SFAMNKSRESYILTLAMPLGSDVGGQTYAWASPRLMEDMPELLDEVVVQEYQRELAEKN------ 668
AURDEDRAFT_176601_AdeI_393226614   425  FYPLVRK--AYAMNRYRESYVAVIAMPVSGDLKGMTYAWGSPRLLKDLLDFLDDCAFEVSHGTMDKRAHGYAKA 496
consensus/100%                       ..shsRK..saAMNp...SYlh.hshP...-h.ub.YsWuSsplbbDh..hlspss.p....ph.c+.......
```

Fig. S12. Multiple sequence alignment of the AdeIX1 domain

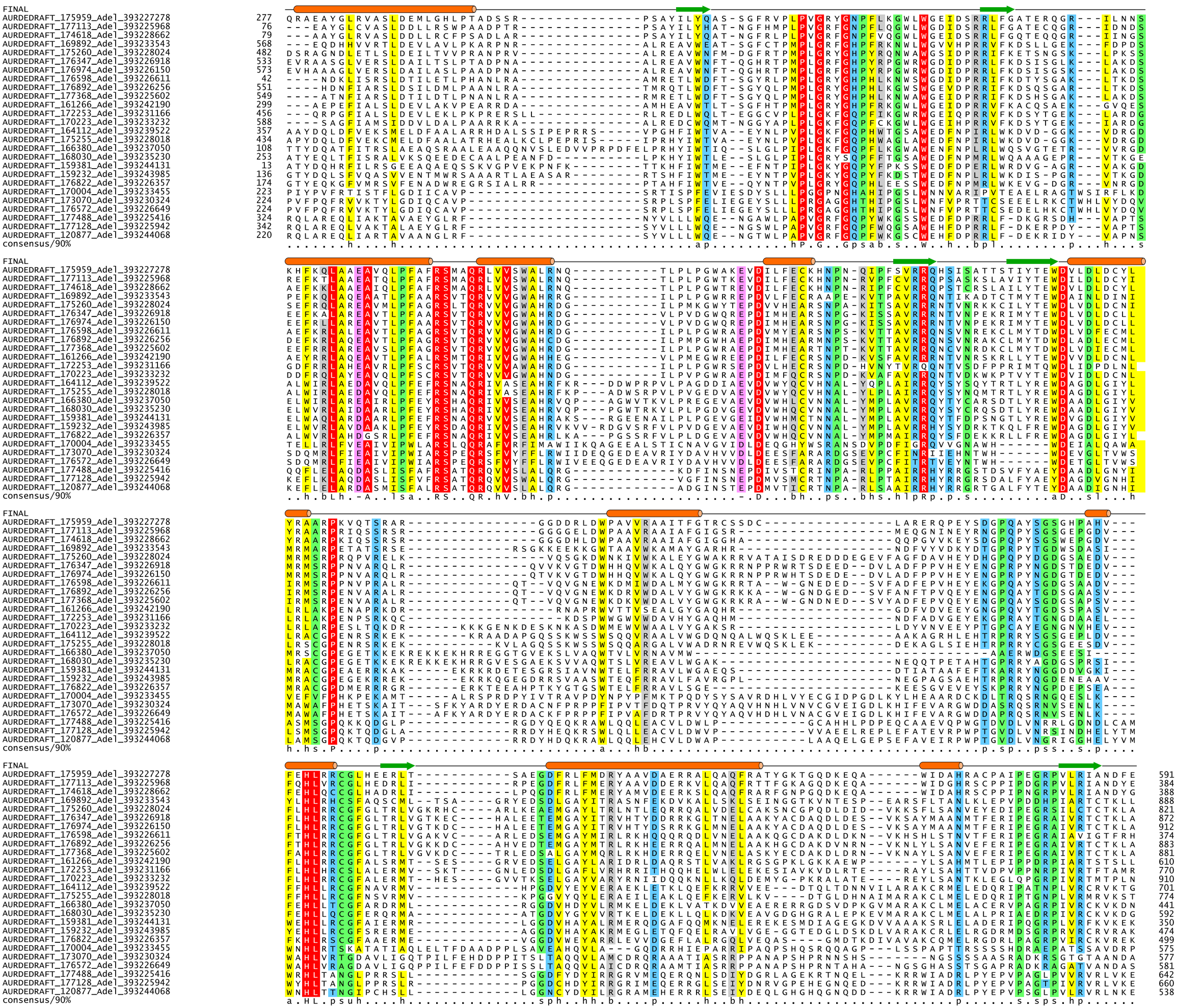

Fig. S13. Multiple sequence alignment of the AdelTA2 domain.

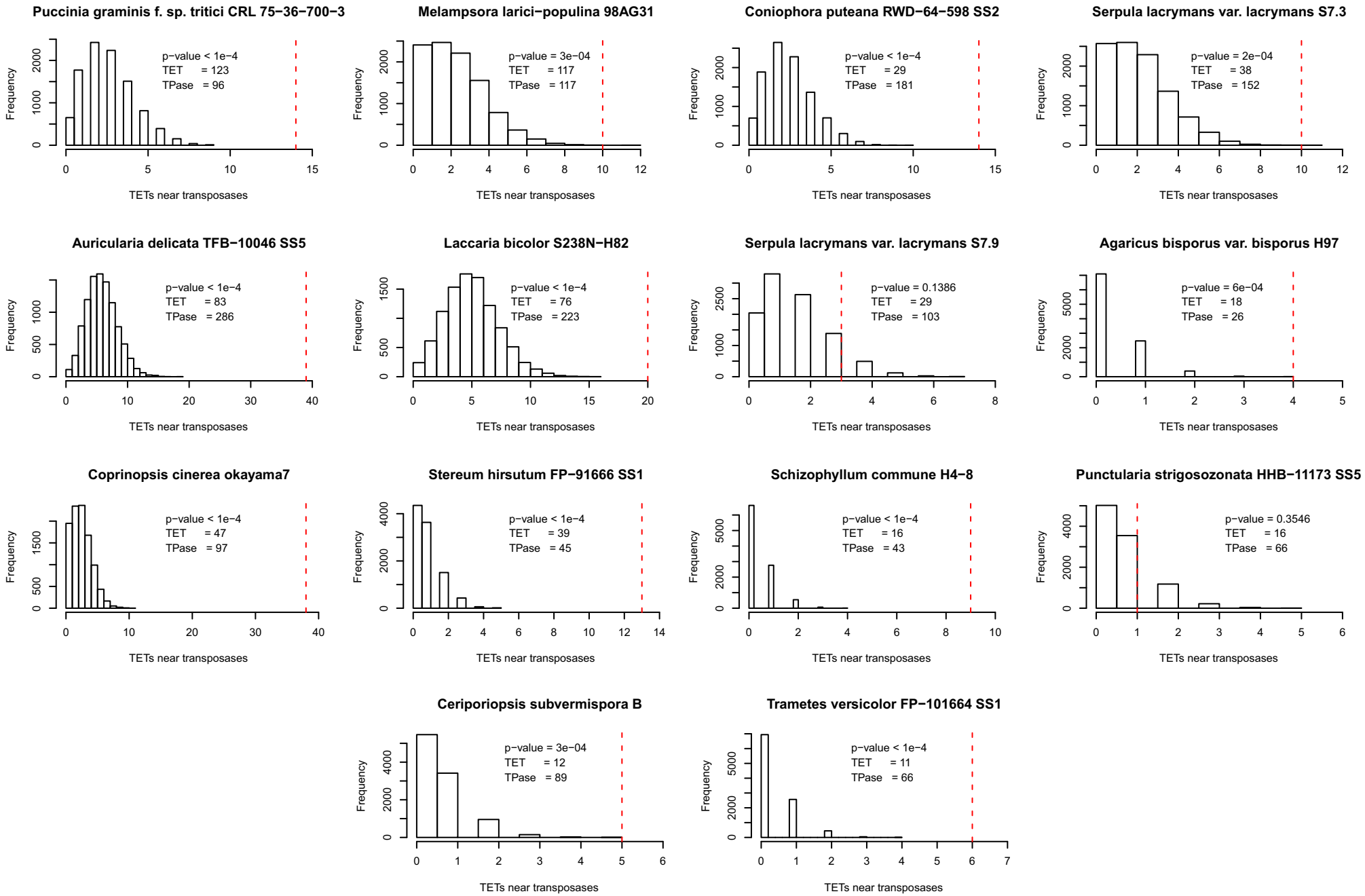

Fig. S14. Simulation of genomic association betweenTET/JBP and transposase genes.
For each fungal species, the graphs show the frequency of randomly placed TET/JBP genes that were found to be located within 10kbp of the nearest randomly placed transposase gene. The total number of TET/JBPs (TET) and transposases (Tpase) identified in each genome is also shown and the red line indicates the observed number of TET/JBP genes near transposases in the real genome. P-values are equal to the total frequency of the number of JBPs near transposases that is equal or higher than the number observed in the real genome.

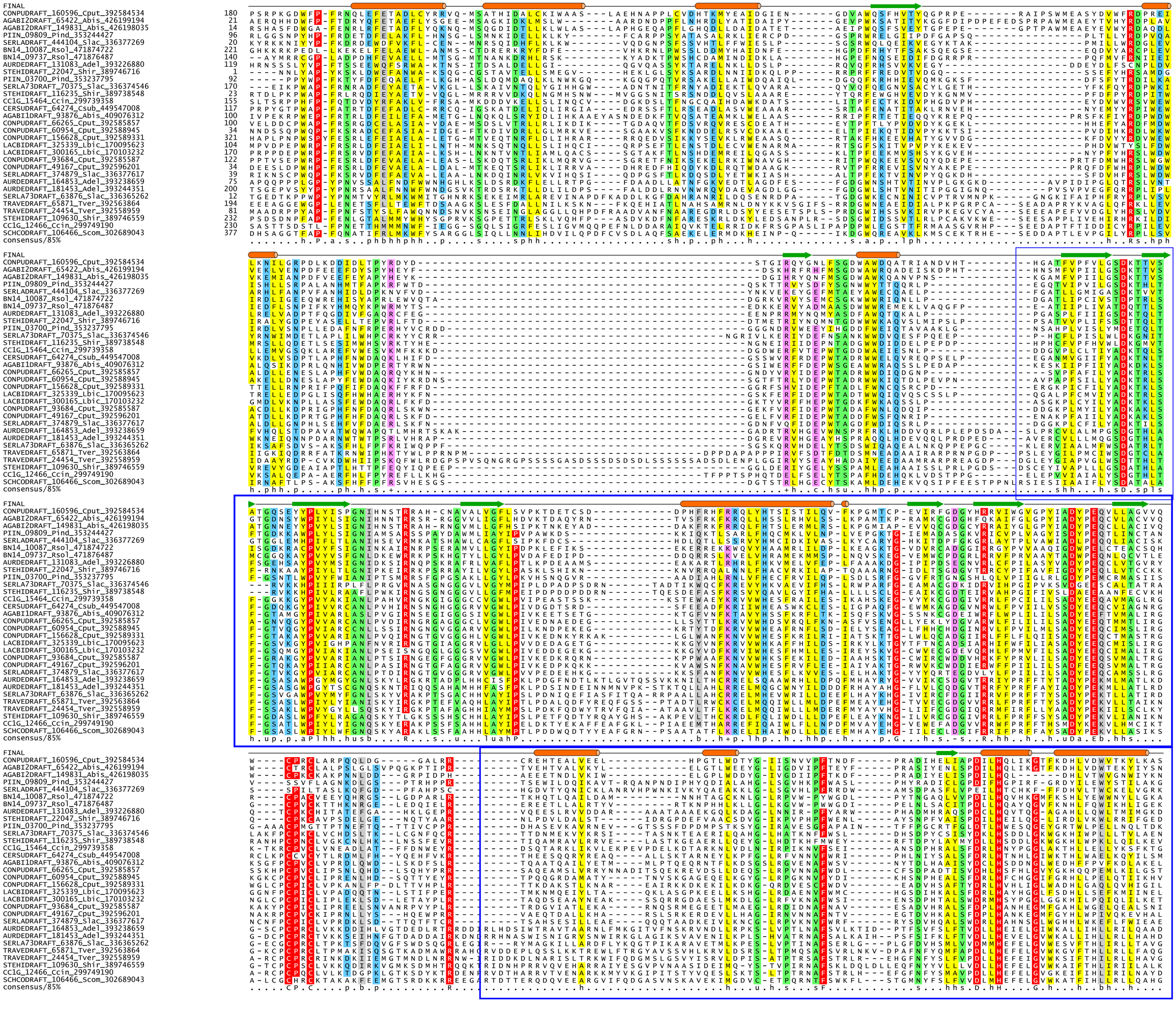

Fig. S15. Multiple sequence alignment of Plavaka transposases. The core of the RNaseH fold is shown in blue boxes.

Fig. S16. Multiple sequence alignment of the BAM/BAH domain.

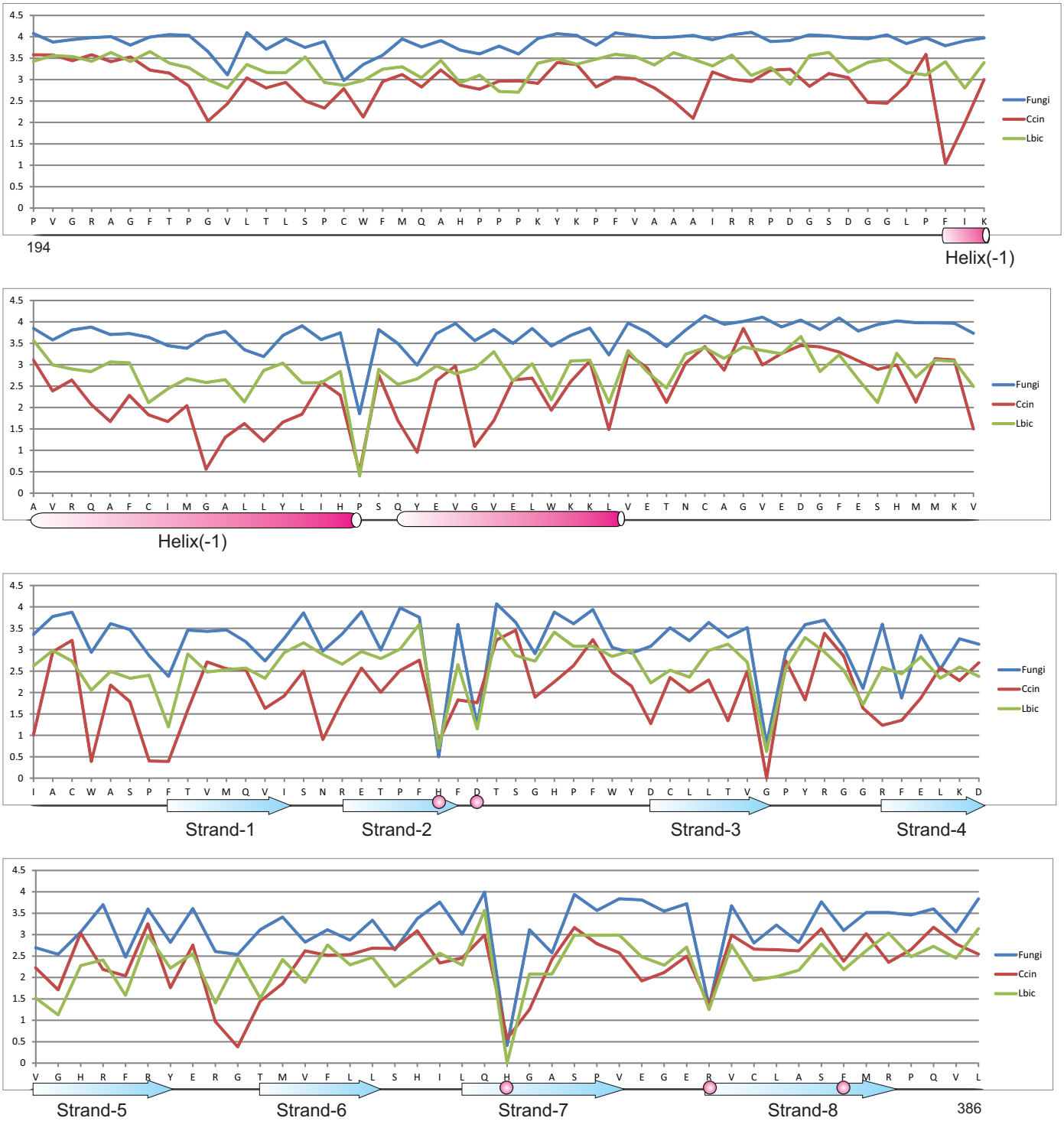

Fig. S17. Shannon entropy plots for Fungal TET/JBP sequences. The entropy plots were derived from a multiple alignments of all Fungal TET/JBP proteins (blue), or from *Coprinopsis cinerea* (red) or *Laccaria bicolor* (green).The TET/JBP sequence of the *Coprinopsis* gene CC1G_05497 (gi: 299740246) is used as a reference sequence (X-axis). For a multiple sequence alignment, the Shannon entropy (H) for each position is given by the equation:

$$H = -\sum_{i=1}^{M} P_i \log_2 P_i$$

where P_i is the fraction of residues of amino acid type *i* and *M* is the number of amino acid types. The Shannon entropy for a residue position ranges from 0 (only one residue at that position) to 4.32 (all 20 residues equally represented at that position). Also shown are the predicted conserved structural elements of the core DSBH of the TET/JBP superfamily. Catalytic residues are marked with a pink circle.

The plot shows that catalytic residues consistently have the lowest entropy suggesting that a majority of fungal TET/JBP sequences are active. Low entropy positions are also observed at the junction of structural elements.The plot shows that the key active site residues of TET/JBP proteins, which chelate the Fe and form a salt bridge with the 2OG, and certain positions critical for the structure of the double stranded beta-helix (DSBH) are the only ones that consistently show a low entropy score (less variability) across all fungal TET/JBP proteins. When this pattern is superimposed on the equivalent plots from just Coprinopsis (Ccin) and Laccaria (Lbic) the same residues show the lowest entropy. The plot also shows that the active site residues of most fungal TET/JBP genes and those in *Coprinopsis* and *Laccaria* are active (Entropy score<1.5). This supports a strong tendency to retain the active site and DSBH stabilizing residues while allowing variability at other positions. This is supportive of selection for active TET/JBPs across the fungal expansion.