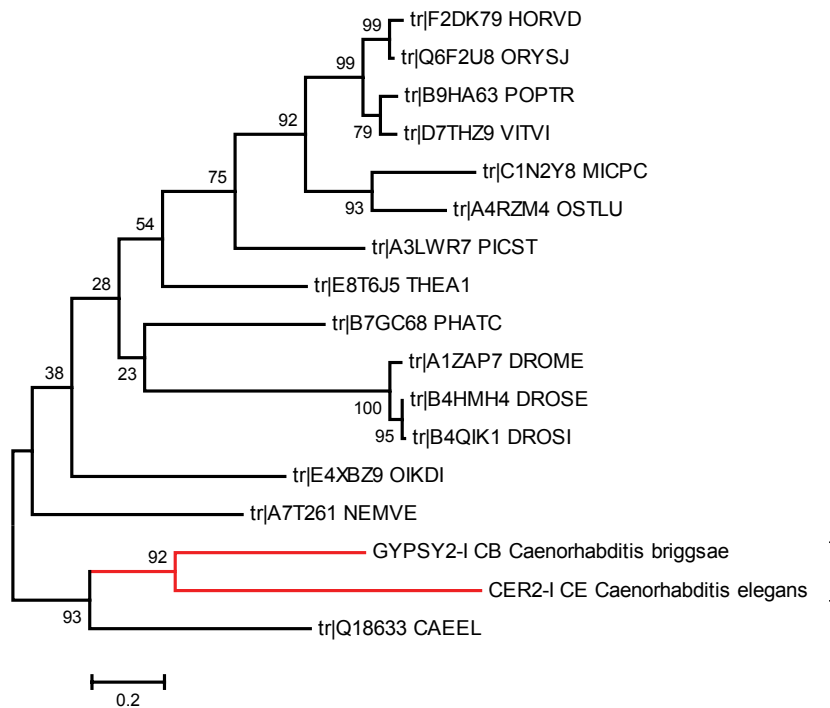


Supplementary figures

Supplementary figure 1. Examples of protein incorporation into TEs.

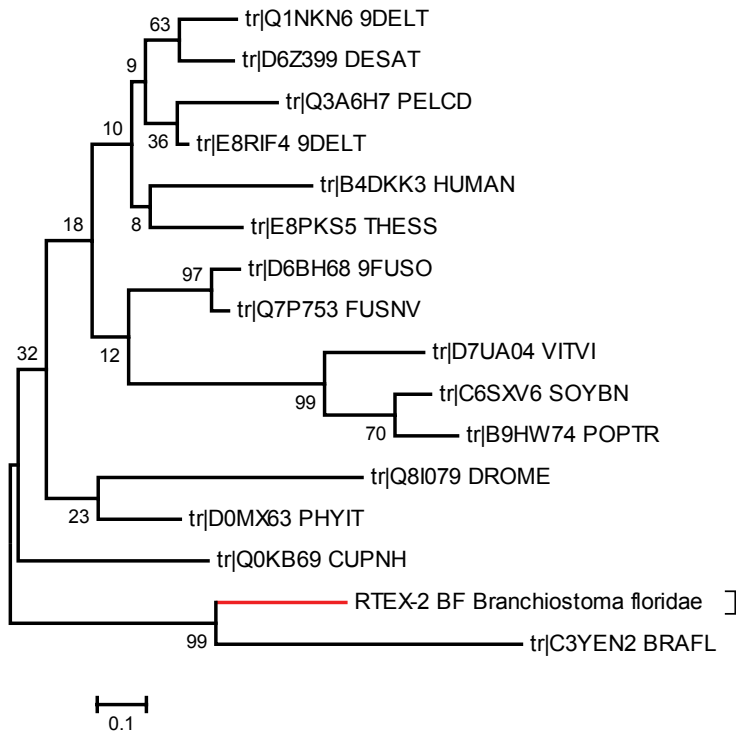
A) Incorporation of an ATPase domain (Pfam PF13087) into LTR retrotransposons in *Caenorhabditis*. The phylogeny shows a maximum likelihood tree (100 bootstrap replicates) of the two retrotransposon families (indicated with red) carrying the domain, and the 15 closest non-TE homologs to GYPSY2-I_CB in Uniprot, (excluding putative/uncharacterized proteins) which include bacteria, plants, fungi and metazoans.



GYPSY2 pol-like protein, the sequence used in the phylogeny is highlighted in red.

```
>GYPSY2-I_CB_ORF, pol-like
MILVSNVLTPLYLFFRQPAQPQFERPQSPDGQVCDYVTLTERLRSRPRINALAHPTEYFSAVVPLKANGINC
WALIDTGASFTVAGQGILHLLGI PRLQEP LSYMAVGLGNNEVKMAGSAVVKFQIGSHVLFQNTHTTGGQC
TPGGPRDYDFIIGNDAHYMTVDVLPSEPAQRISESTTLQMIQDFRNETTKWMTIDTPITSVPNGKRNP
VATLSRFDCDRIFALYIVKESAGGNLHLNPAHRELVS YDRPDLYTVHLTRHSEDI LNPGPPNPANYIIGD
LLVVL SLKRRPGAKLNEIKRDFVDILDMNKCHFQVQVNFALVQRKVEKQPLWKMDLKEPIDPRRKLKM
PWKAECIGYNQNVVEAKIALFKELHPQDKCTAADILVVPVLEPASTLRSLLPGDARHAVITATNVQPPI
VPLKPTVTAVRVLPGKFRKYIS IAPNAFKGAESDPDYDLGGKALLWTFGLAATLAIANKGRDTTTTYST
TITESETTGRRTVTFITIRRPSTAAPVSSWTRSTAFKMDISPGGTCDEVESATLINTQLTIRAKFTFES
QSRINEKNLLNQNAV VVQQLDDNLHQLRTFPEPEDWLKVPVDS PGNVLIALTGGPEIDGIMPWKD TDVV
DFAGSAVKLSEEQSYANTLAKTKIAGVIANAPPGVGKTLMVAGAAIHA IRETNDLRS IQTLTGMTNVSV
LPMVKTLRLTDPETRYPPVRLISESHREHLDPALRTEIDYPILWIKVFKDLLHRIDPSTAVATIRENPD
MSAIQHLHAKKIDPEEF AHHFIRKAVAKGPNDRNTTPKAELFRKLYNPRIVVGTVASVLESFKTPDWTW
LRENVSAIQVDEASQLPLYALALALMFQRRARLALVGDHQLPPFEDSDLPKDIAAFATSRVLQEAIKRR
TLPIIDLTVGRRCPQVTRMYNELFYNQRLTSLWSPEKEAELRAFTAALKFGNDFPIQLDLQSSHTQSG
TSLTNIEEAKAAVRIARTIQTRVGTNDIGILCFYKAQAGEVAALLGDAPFYVGTIDGSQGHEFQAVIILT
TRTSSFRDSEFIEDCRRINVAISRTRKRICCIIVDRPKVTTNGTWSRLIRRI PHEAKSNFSVMTARWL
```

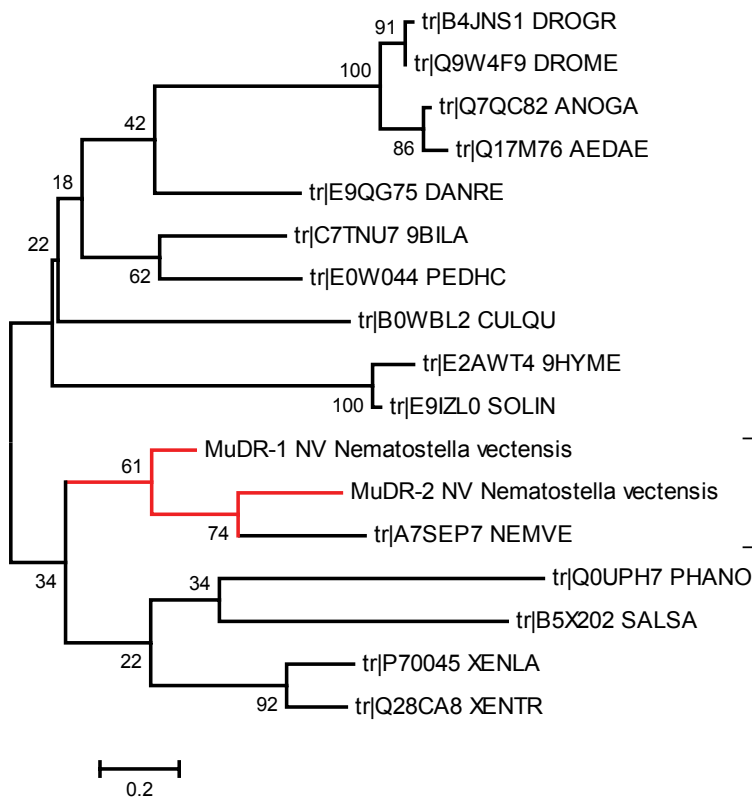
B) Incorporation of a DnaJ chaperone domain (Pfam PF00226) into the ORF1 protein of a non-LTR retrotransposon (RTEX-2_BF) of *Branchiostoma floridae*. The DnaJ domain is distributed widely across eukaryotes and prokaryotes, but the TEs carrying the protein fragment are present only in *Branchiostoma floridae*. The tree shows the domain from the RTEX-2_BF retrotransposon and its 15 closest homologs from Uniprot. The incorporation of this domain has already been noted in RepBase.



ORF1 protein of RTEX-2_BF non-LTR retrotransposon, the region highlighted in red was used in the phylogeny.

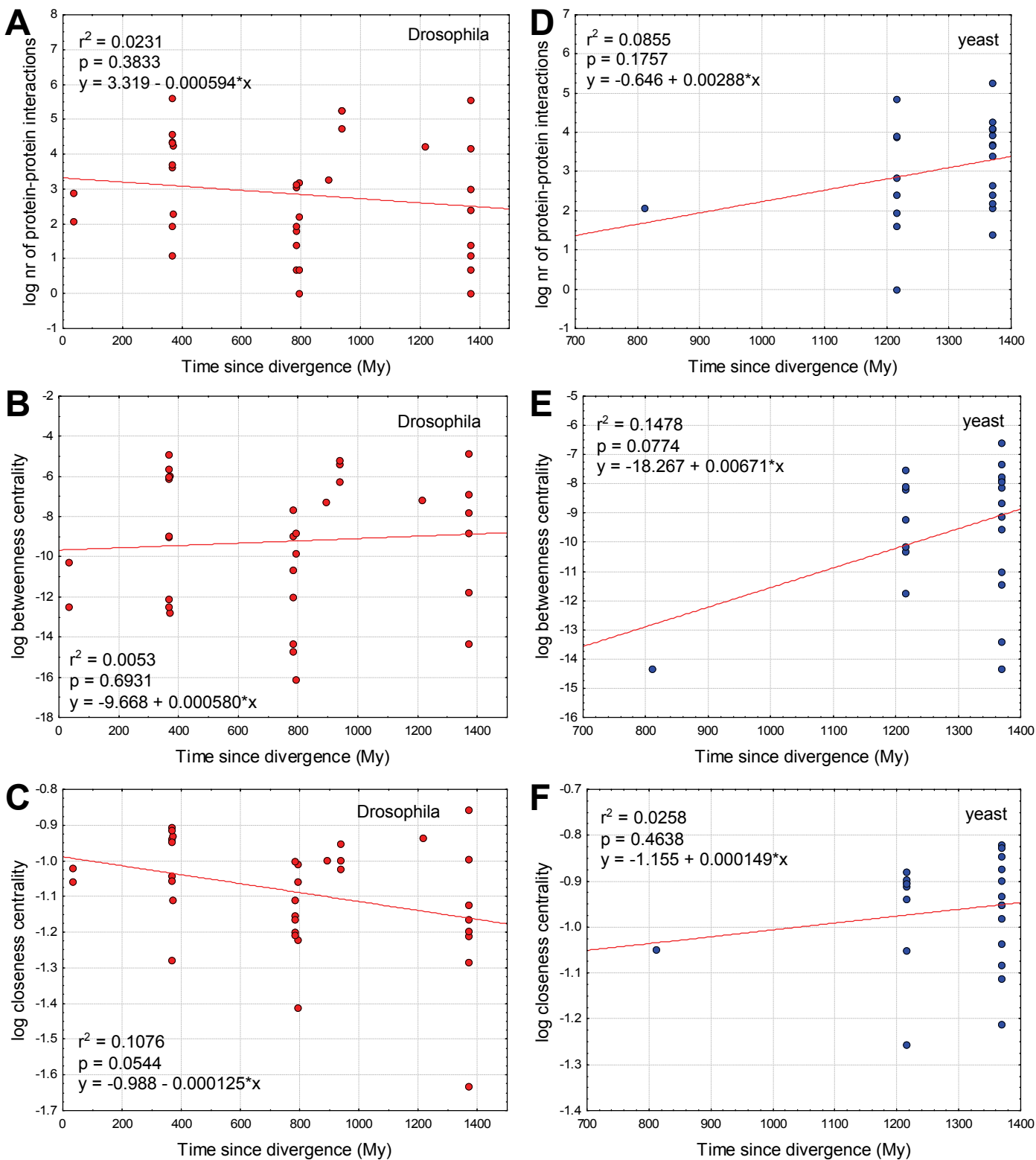
```
>RTEX-2_BF_1p
MADNSEELYRTLGVASNASKEDITEAYKKEVANYENASKLGRKEPLFKEADTRYREVS KAFVVLADAARR
QQYDAGLPASNSTKTKV KISKERTDYQVQINKQSVTIYLPNHLTNPWLKTC EEHYRTLATVGDKEENGHQI
KTSFVDTRSGQPVGVS VSIKFFETTQKLLIQGSAYLLWFAEDYPELKKTVTSLSPNNSNGNAHIQASMPDP
DANSTPASPASVAVNLP PCPTCQEPVESD ICACNTVPKALLDVNSNNVNGTESQTGPTGDHDEFNMANTS
VTPDYSIVQESMTKLET CLANCITDKKLFEDSVMHKLSAIEARVKS CERPHTCDGFTTEEKERLSSEISR
LDKEKRELELKVKSLE QRHERLSTAVRKA EATNLKRQTSTVETQTSESNSDRAERLLHEATISVHNRQV
LDSNDGISHG SVKHSNNNESSCLPRAHSQPNQSPRETRRTHCESNQPPCVTRDRKSSQHD TMANNQHIK
ARSNQQNND AQLDLL IILGDSNTRPLKTDILYPNKRVTKELTFNLREAI DYIQTSTLPDPKTI LFHVG TND
VRDARDPTTVSEGFREL VQITHDKYPQSHIVLSPILPRDDPNLQAIGDNVNAFLKVVAD ETSYVHIIDNS
NFSYSGTTSIKKSLYNS DGYHLNRNGTRVLAANI KRTVNSLIGLQYLSRGSQPVACDQFATS PKRTTPN
RSYRDAVLGAPAGG PPEPSTRSCPPPH TQPARKQEHS SAPSQDRRSSNHTSSDDPVNGSSDRCPAPDRRP
PPFGLHPPVGPWGP PFPAGPWGPP PPA GPWGP PPA GPWGP PPA GPWGP PPA GPWGP PPA GPWGP
SPFFPSPSHLLGKSGRFPDSYPCSAQNSKQQFP PAWLWQPGTPWRGPPSMW
```

C) Incorporation of a C2H2 type Zinc finger domain (Pfam PF00096) into MuDR DNA transposons in *Nematostella vectensis*. Its non-TE homologs are present across all major branches of eukaryotes, but in TEs it is restricted to MuDR DNA transposon families of *Nematostella*. The maximum likelihood tree (100 bootstrap replicates) shows the phylogeny of domains from MuDR transposons and their 15 closest homologs from Uniprot, excluding several highly similar sequences from *Drosophila*. Interestingly, the domains that were picked up by TEs were later reused in a protein of unknown function, A7SEP7.



The sequence that was picked up by the TEs is not part of an annotated TE ORF (although it is subject to significant purifying selection):

```
>MuDR-1_NV Nematostella_vectensis, frame 5, 355-444
AQCQREELDMGIAEIEEKPCSCNTTFKRRKDMKYHKKGCHECICEICHKQFNHEQKLKRHKNEHNATYKC
GTCQKSFÆKRNLKRHESTH
```



Supplementary figure 2. Relationships between the time of divergence of the TE hosts from *Drosophila melanogaster*/yeast, and the protein-protein interaction network parameters (number of interactions, betweenness centrality, closeness centrality) of the proteins incorporated by the TEs in the hosts. None of the regressions are significant, indicating that the divergence time introduces no measurement biases for these network parameters (although in yeast the analysis has little power due to the narrow span of divergence times). A-C) *Drosophila*; D-F) yeast.