Additional File 1

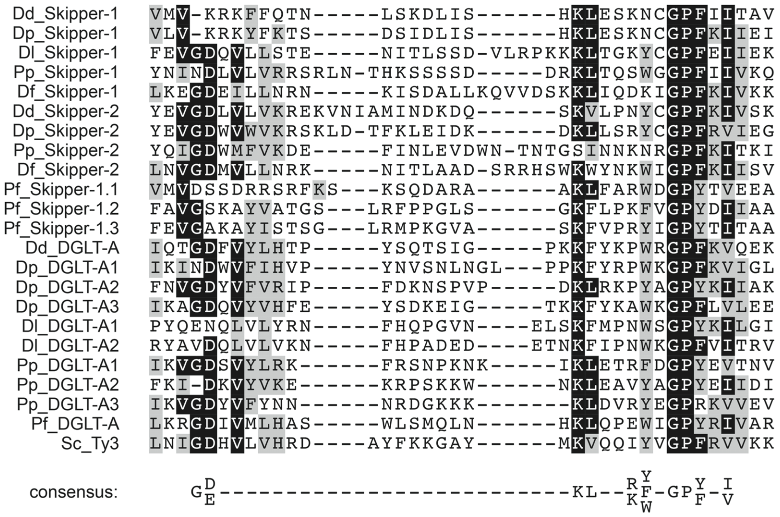
Convergent evolution of tRNA gene targeting preferences in compact genomes

Thomas Spaller, Eva Kling, Gernot Glöckner, Falk Hillmann, Thomas Winckler

Content

Figure S1: Alignments of GPY/F motifs in the carboxy-terminal regions of integrase domains in dictyostelid LTR retrotransposons.

Alignments were generated with CLUSTALX [1] and BOXSHADE [2]. Black boxes indicate invariant amino acids and gray boxes represent similar amino acids. Shading is to a 50% consensus. Ty3 (GenBank M23367) is shown for comparison.The consensus of this motif as suggested by Malik and Eickbush [3] is indicated. Abbreviations: Dd: *Dictyostelium discoideum*; Dp: *D. purpureum*; Dl: *D. lacteum*; Pp: *Polysphondylium pallidum*; Df: *D. fasciculatum*; Pf: *Protostelium fungivorum*; Sc: *Saccharomyces cerevisiae.*



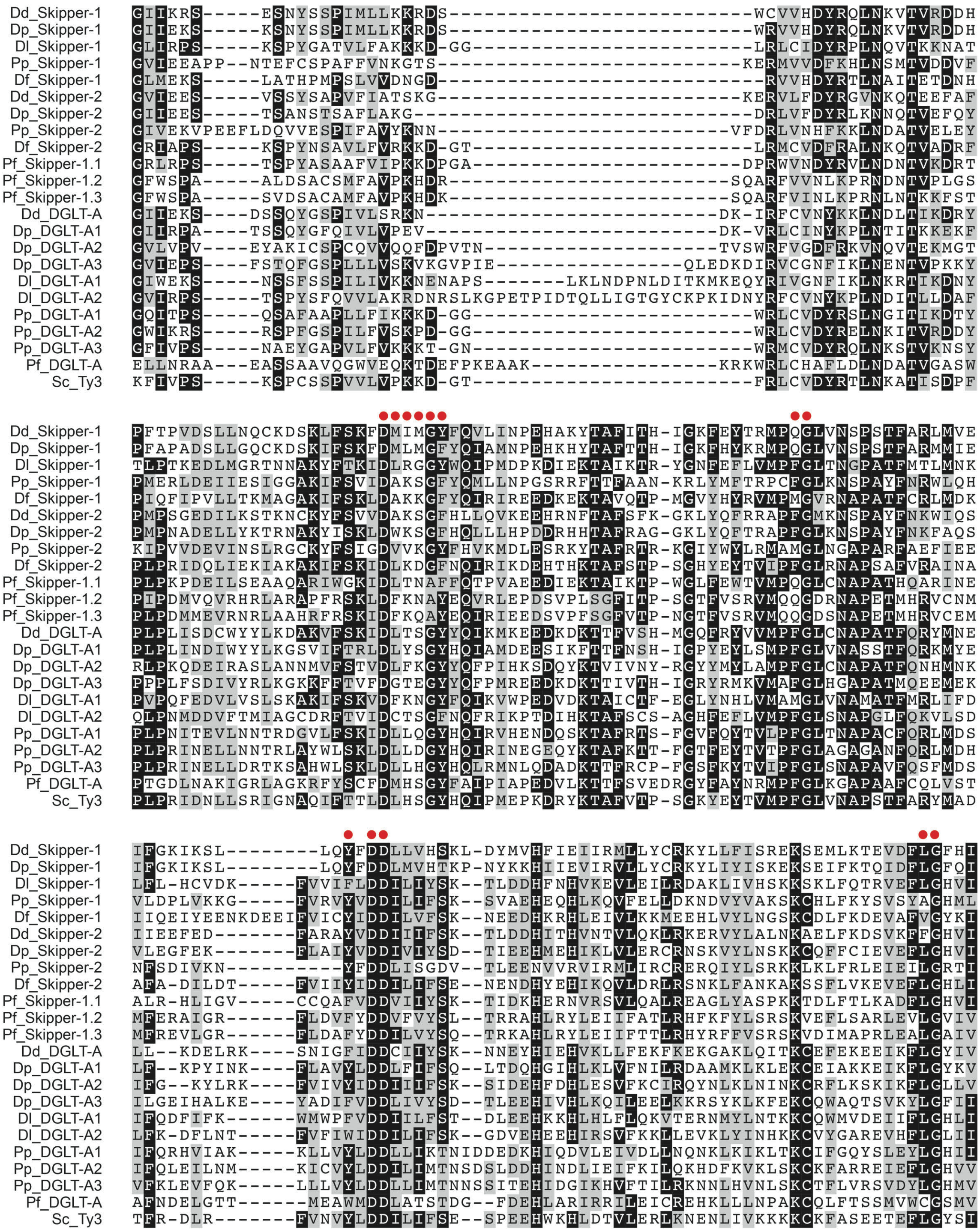
```
Dd_Skipper-1    VMV-KRKFFQTN-----LSKDLIS------HKLESKNCGPFIITAV
Dp_Skipper-1    VLV-KRKYFKTS-----DSIDLIS------HKLESKNCGPFKIIEI
Dl_Skipper-1    FEVGDQVLLSTE-----NITLSSD-VLRPKKKLTGKYCGPFEIIEK
Pp_Skipper-1    YNINDLVLVRRSRLN-THKSSSSD-----DRKLTQSWGGPFIIVKQ
Df_Skipper-1    LKEGDEILLNRN-----KISDALLKQVVDSKKLIQDKIGPFKIVKK
Dd_Skipper-2    YEVGDLVLVKREKVNIAMINDKDQ------SKVLPNYCGPFKIVSK
Dp_Skipper-2    YEVGDWVWVKRSKLD-TFKLEIDK------DKLLSRYCGPFRVIEG
Pp_Skipper-2    YQIGDWMFVKDE-----FINLEVDWN-TNTGSINNKNRGPFKITKI
Df_Skipper-2    LNVGDMVLLNRK-----NITLAAD-SRRHSWKWYNKWIGPFKIISV
Pf_Skipper-1.1  VMVDSSDRRSRFKS---KSQDARA------AKLFARWDGPYTVEEA
Pf_Skipper-1.2  FAVGSKAYVATGS---LRFPPGLS------GKFLPKFVGPYDIIAA
Pf_Skipper-1.3  FEVGAKAYISTSG---LRMPKGVA------SKFVPRYIGPYTITAA
Dd_DGLT-A       IQTGDFVYLHTP-----YSQTSIG-----PKKFYKPWRGPFKVQEK
Dp_DGLT-A1      IKINDWVFIHVP-----YNVSNLNGL---PPKFYRPWKGPFKVIGL
Dp_DGLT-A2      FNVGDYVFVRIP-----FDKNSPVP-----DKLRKPYAGPYKIIAK
Dp_DGLT-A3      IKAGDQVYVHFE-----YSDKEIG-----TKKFYKAWKGPFLVLEE
Dl_DGLT-A1      PYQENQLVLYRN-----FHQPGVN----ELSKFMPNWSGPYKILGI
Dl_DGLT-A2      RYAVDQLVLVKN-----FHPADED----ETNKFIPNWKGPFVITRV
Pp_DGLT-A1      IKVGDSVYLRK------FRSNPKNK-----IKLETRFDGPYEVTNV
Pp_DGLT-A2      FKI-DKVYVKE------KRPSKKW------NKLEAVYAGPYEIIDI
Pp_DGLT-A3      IKVGDYVFYNN------NRDGKKK-------KLDVRYEGPRKVVEV
Pf_DGLT-A       LKRGDIVMLHAS-----WLSMQLN------HKLQPEWIGPYRIVAR
Sc_Ty3          LNIGDHVLVHRD----AYFKKGAY------MKVQQIYVGPFRVVKK

                    D                             RY   Y I
consensus:         G -------------------------KL-- F-GP - -
                    E                             KW   F V
```

Figure S2: Alignments of RT, RNH, and IN domains of dictyostelid LTR retrotransposons.

Core domains of reverse transcriptase (RT) (**A**), ribonuclease H (RNH) (**B**), and integrase (IN) (**C**) were determined by searching the Conserved Domain Database [4]. Alignments were generated with CLUSTALX [1]. Shading is to a 50% consensus and was generated with BOXSHADE [2]. Black boxes indicate invariant amino acids and gray boxes represent similar amino acids. Active site residues are indicated by red dots. The sequence of Ty3 (GenBank M23367) was aligned for comparison. Abbreviations: Dd: *Dictyostelium discoideum*; Dp: *D. purpureum*; Dl: *D. lacteum*; Pp: *Polysphondylium pallidum*; Df: *D. fasciculatum*; Pf: *Protostelium fungivorum*; Sc: *Saccharomyces cerevisiae.*

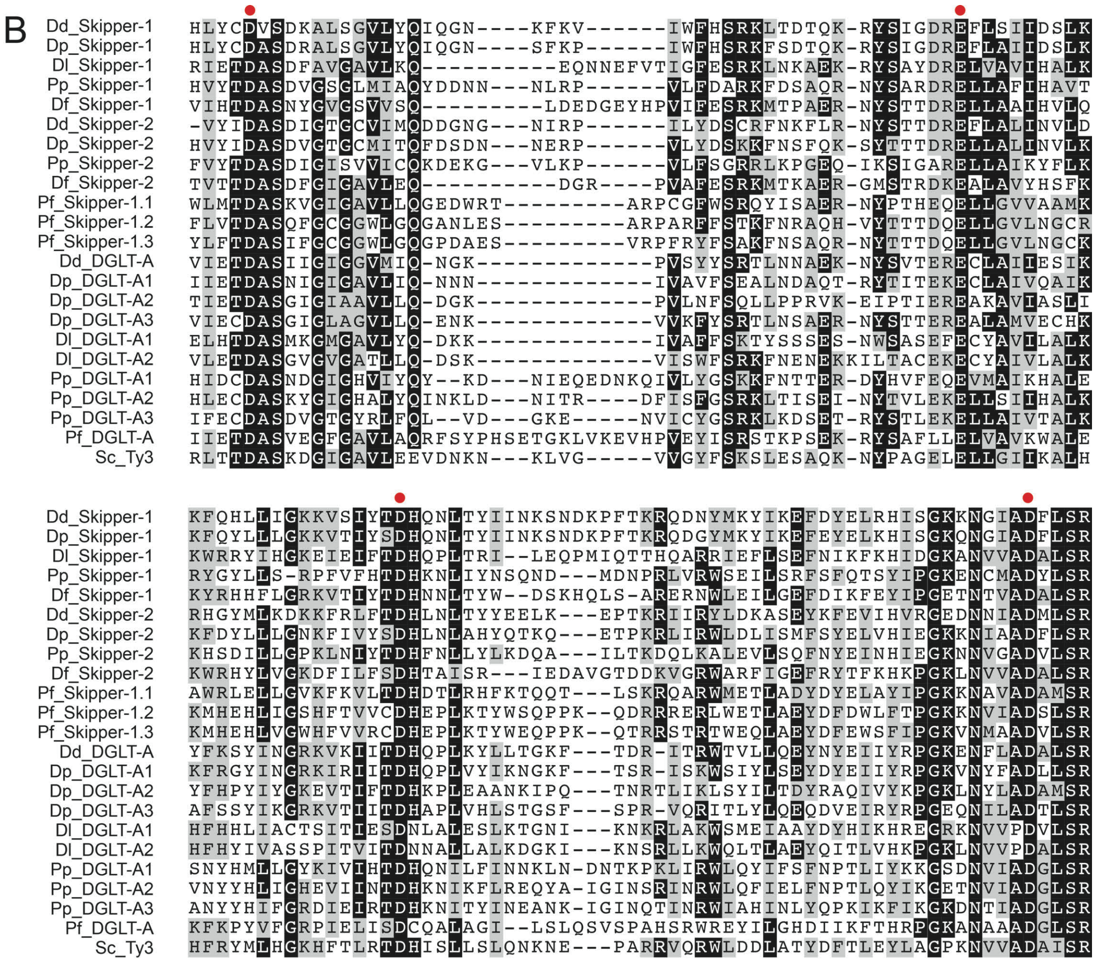
A



```
  Dd_Skipper-1  GIIKRS-----ESNYSSPIMLLKKRDS--------------------------WCVVHDYRQLNKVTVRDDH
  Dp_Skipper-1  GIIEKS-----KSNYSSPIMLLKKRDS--------------------------WRVVHDYRQLNKVTVRDDH
  Dl_Skipper-1  GLIRPS-----KSPYGATVLFAKKKD-GG------------------------LRLCIDYRPLNQVTKKNAT
  Pp_Skipper-1  GVIEEAPP--NTEFCSPAFFVNKGTS--------------------------KERMVVDFKHLNSMTVDDVF
  Df_Skipper-1  GLMEKS-----LATHPMPSLVVDNGD----------------------------RVVHDYRTLNAITETDNH
  Dd_Skipper-2  GVIEES-----VSSYSAPVFIATSKG--------------------------KERVLFDYRGVNKQTEEFAF
  Dp_Skipper-2  GIIEES-----TSANSTSAFLAKG-----------------------------DRLVFDYRLKNNQTVEFQY
  Pp_Skipper-2  GIVEKVPEEFLDQVVESPIFAVYKNN-------------------------VFDRLVNHFKKLNDATVELEY
  Df_Skipper-2  GRIAPS-----KSPYNSAVLFVRKKD-GT------------------------LRMCVDFRALNKQTVADRF
Pf_Skipper-1.1  GRLRPS-----TSPYASAAFVIPKKDPGA-----------------------DPRWVNDYRVLNDNTVKDRT
Pf_Skipper-1.2  GFWSPA-----ALDSACSMFAVPKHDR------------------------SQARFVVNLKPRNDNTVPLGS
Pf_Skipper-1.3  GFWSPA-----SVDSACAMFAVPKHDK------------------------SQARFVINLKPRNLNTKKFST
     Dd_DGLT-A  GIIEKS----DSSQYGSPIVLSRKN-------------------------DK-IRFCVNYKKLNDLTIKDRY
    Dp_DGLT-A1  GIIRPA----TSSQYGFQIVLVPEV-------------------------DK-VRLCINYKPLNTIITKKEKF
    Dp_DGLT-A2  GVLVPV----EYAKICSPCQVVQQFDPVTN--------------------TVSWRFVGDFRKVNQVTEKMGT
    Dp_DGLT-A3  GVIEPS----FSTQFGSPLLLVSKVKGVPIE----------------QLEDKDIRVCGNFIKLNENTVPKKY
    Dl_DGLT-A1  GIWEKS-----NSSFSSPILIVKKNENAPS-------LKLNDPNLDITKMKEQYRIVGNFIKLNKRTIKDNY
    Dl_DGLT-A2  GVIRPS-----TSPYSFQVVLAKRDNRSLKGPETPIDTQLLIGTGYCKPKIDNYRFCVNYKPLNDITLLDAF
    Pp_DGLT-A1  GQITPS-----QSAFAAPLLFIKKKD-GG------------------------WRLCVDYRSLNGITIKDTY
    Pp_DGLT-A2  GWIKRS-----RSPFGSPILFVSKPD-GG------------------------WRLCVDYRELNKITVRDDY
    Pp_DGLT-A3  GFIVPS-----NAEYGAPVLFVKKKT-GN------------------------WRMCVDYRQLNKSTVKNSY
     Pf_DGLT-A  ELLNRAA---EASSAAVQGWVEQKTDEFPKEAAK----------------KRKWRLCHAFLDLNDATVGASW
        Sc_Ty3  KFIVPS-----KSPCSSPVVLVPKKD-GT------------------------FRLCVDYRTLNKATISDPF

                                     ●●●●●●                             ●●
  Dd_Skipper-1  PFTPVDSLLNQCKDSKLFSKFDMIMGYFQVLINPEHAKYTAFITH-IGKFEYTRMPQGLVNSPSTFARLMVE
  Dp_Skipper-1  PFAPADSLLGQCKDSKIFSKFDMLMGFYQIAMNPEHKHYTAFTTH-IGKFHYKRMPQGLVNSPSTFARMMIE
  Dl_Skipper-1  TLPTKEDLMGRTNNAKYFTKIDLRGGYWQIPMDPKDIEKTAIKTR-YGNFEFLVMPFGLTNGPATFMTLMNK
  Pp_Skipper-1  PMERLDEIIESIGGAKIFSVIDAKSGFYQMLLNPGSRRFTTFAAN-KRLYMFTRPCFGLKNSPAYFNRWLQH
  Df_Skipper-1  PIQFIPVLLTKMAGAKIFSKLDAKKGFYQIRIREEDKEKTAVQTP-MGVYHYRVMPMGVRNAPATFCRLMDK
  Dd_Skipper-2  PMPSGEDILKSTKNCKYFSVVDAKSGFHLLQVKEEHRNFTAFSFK-GKLYQFRRAPFGMKNSPAYFNKWIQS
  Dp_Skipper-2  PMPNADELLYKTRNAKYISKLDWKSGFHQLLLHPDDRHHTAFRAG-GKLYQFTRAPFGLKNSPAYFNKFAQS
  Pp_Skipper-2  KIPVVDEVINSLRGCKYFSIGDVVKGYFHVKMDLESRKYTAFRTR-KGIYWYLRMAMGLNGAPARFAEFIEE
  Df_Skipper-2  PLPRIDQLIEKIAKAKIFSKIDLKDGFNQIRIKDEHTHKTAFSTP-SGHYEYTVIPFGLRNAPSAFVRAINA
Pf_Skipper-1.1  PLPKPDEILSEAAQARIWGKIDLTNAFFQTPVAEEDIEKTAIKTP-WGLFEWTVMPQGLCNAPATHQARINE
Pf_Skipper-1.2  PIPDMVQVRHRLARAPFRSKLDFKNAYEQVRLEPDSVPLSGFITP-SGTFVSRVMQQGDRNAPETMHRVCNM
Pf_Skipper-1.3  PLPDMMEVRNRLAAHRFRSKIDFKQAYEQIRIEEDSVPFSGFVTP-NGTFVSRVMQQGDSNAPETMHRVCEM
     Dd_DGLT-A  PLPLISDCWYYLKDAKVFSKIDLTSGYYQIKMKEEDKDKTTFVSH-MGQFRYVVMPFGLCNAPATFQRYMNE
    Dp_DGLT-A1  PLPLINDIWYYLKGSVIFTRLDLYSGYHQIAMDEESIKFTTFNSH-IGPYEYLSMPFGLVNASSTFQRKMYE
    Dp_DGLT-A2  RLPKQDEIRASLANNMVFSTVDLFKGYYQFPIHKSDQYKTVIVNY-RGYYMYLAMPFGLCNAPATFQNHMNK
    Dp_DGLT-A3  PPPLFSDIVYRLKGKKFFTVFDGTEGYYQFPMREEDKDKTTIVTH-IGRYRMKVMAFGLHGAPATMQEEMEK
    Dl_DGLT-A1  PVPQFEDVLVSLSKAKIFSKVDFKNGYFQIKVWPEDVDKTAICTF-EGLYNHLVMAMGLVNAMATFMRLIFD
    Dl_DGLT-A2  QLPNMDDVFTMIAGCDRFTVIDCTSGFNQFRIKPTDIHKTAFSCS-AGHFEFLVMPFGLSNAPGLFQKVLSD
    Pp_DGLT-A1  PLPNITEVLNNTRDGVLFSKIDLLQGYHQIRVHENDQSKTAFRTS-FGVFQYTVLPFGLTNAPACFQRLMDS
    Pp_DGLT-A2  PLPRINELLNNTRLAYWLSKLDLLDGYHQIRINEGEQYKTAFKTT-FGTFEYTVTPFGLAGAGANFQRLMDH
    Pp_DGLT-A3  PLPRINELLDRTKSAHWLSKLDLLHGYHQLRMNLQDADKTTFRCP-FGSFKYTVIPFGLSNAPAVFQSFMDS
     Pf_DGLT-A  PTGDLNAKIGRLAGKRFYSCFDMHSGYFAIPIAPEDVLKTTFSVEDRGYFAYNRMPFGLKGAPAAFCQLVST
        Sc_Ty3  PLPRIDNLLSRIGNAQIFTTLDLHSGYHQIPMEPKDRYKTAFVTP-SGKYEYTVMPFGLVNAPSTFARYMAD

                                  ● ●●                                             ●●
  Dd_Skipper-1  IFGKIKSL--------LQYFDDLLVHSKL-DYMVHFIEIIRMLLYCRKYLLFISREKSEMLKTEVDFLGFHI
  Dp_Skipper-1  IFGKIKSL--------LQYFDDLMVHTKP-NYKKHIIEVIRVLLYCRKYLIYISKEKSEIFKTQIDFLGFQI
  Dl_Skipper-1  LFL-HCVDK-----FVVIFLDDILIYSK--TLDDHFNHVKEVLEILRDAKLIVHSKKSKLFQTRVEFLGHVI
  Pp_Skipper-1  VLDPLVKKG-----FVRVYVDDILIFSK--SVAEHEQHLKQVFELLDKNDVYYAKSKCHLFKYSVSYAGHML
  Df_Skipper-1  IIQEIYEENKDEEIFVICYIDDILVFSK--NEEDHKRHLEIVLKKMEEHLVYLNGSKCDLFKDEVAFVGYKI
  Dd_Skipper-2  IIEEFED-------FARAYVDDIIIFSK--TLDDHITHVNTVLQKLRKERVYLALNKAELFKDSVKFFGHVI
  Dp_Skipper-2  VLEGFEK-------FLAIYVDDIVIYSD--TIEEHMEHIKLVLERCRNSKVYLNSKKCQFFCIEVEFLGHVI
  Pp_Skipper-2  NFSDIVKN----------YFDDLISGDV--TLEENVVRVIRMLIRCRERQIYLSRKKLKLFRLEIEILGRTI
  Df_Skipper-2  AFA-DILDT-----FVIIYIDDILIFSE--NENDHYEHIKQVLDRLRSNKLFANKAKSSFLVKEVEFLGHLI
Pf_Skipper-1.1  ALR-HLIGV-----CCQAFVDDVIIYSK--TIDKHERNVRSVLQALREAGLYASPKKTDLFTLKADFLGHVI
Pf_Skipper-1.2  MFERAIGR------FLDVFYDDVFVYSL--TRRAHLRYLEIIFATLRHFKFYLSRSKVEFLSARLEVLGVIV
Pf_Skipper-1.3  MFREVLGR------FLDAFYDDILVYSQ--TRKAHLRYLEIIFTTLRHYRFFVSRSKVDIMAPRLEALGAIV
     Dd_DGLT-A  LL--KDELRK----SNIGFIDDCIIYSK--NNEYHIEHVKLLFEKFKEKGAKLQITKCEFEKEEIKFLGYIV
    Dp_DGLT-A1  LF--KPYINK----FLAVYLDDLFIFSQ--LTDQHGIHLKLVFNILRDAAMKLKLEKCEIAKKEIKFLGYKV
    Dp_DGLT-A2  IFG--KYLRK----FVIVYIDDIIIFSK--SIDEHFDHLESVFKCIRQYNLKLNINKCRFLKSKIKFLGLLV
    Dp_DGLT-A3  ILGEIHALKE----YADIFVDDLIVYSD--TLEEHIVHLKQILEELKKRSYKLKFEKCQWAQTSVKYLGFII
    Dl_DGLT-A1  IFQDFIFK------WMWPFVDDIILFST--DLEEHKKHLHLFLQKVTERNMYLNTKKCQWMVDEIIFLGHLI
    Dl_DGLT-A2  LFK-DFLNT-----FVFIWIDDILIFSK--GDVEHEEHIRSVFKKLLEVKLYINHKKCVYGAREVHFLGLII
    Pp_DGLT-A1  IFQRHVIAK-----KLLVYLDDLLIKTNIDDEDKHIQDVLEIVDLLNQNKLKIKLTKCIFGQYSLEYLGHII
    Pp_DGLT-A2  IFQLEILNM-----KICVYLDDILIMTNSDSLDDHINDLIEIFKILQKHDFKVKLSKCKFARREIEFLGHVV
    Pp_DGLT-A3  VFKLEVFQK-----LLLVYLDDLLIMTNNSSITEHIDGIKHVFTILRKNNLHVNLSKCTFLVRSVDYLGHMV
     Pf_DGLT-A  AFNDELGTT------MEAWMDDLATSTDG--FDEHLARIRRILEICREHKLLLNPAKCQLFTSSMVWCGSMV
        Sc_Ty3  TFR-DLR-------FVNVYLDDILIFSE--SPEEHWKHLDTVLERLKNENLIVKKKKCKFASEETEFLGYSI
```


B
Dd_Skipper-1 HLYCDVSDKALSGVLYQIQGN----KFKV-------IWFHSRKLTDTQK-RYSIGDREFLSIIDSLK
Dp_Skipper-1 HLYCDASDRALSGVLYQIQGN----SFKP-------IWFHSRKFSDTQK-RYSIGDREFLAIIDSLK
Dl_Skipper-1 RIETDASDFAVGAVLKQ-----------EQNNEFVTIGFESRKLNKAEK-RYSAYDRELVAVIHALK
Pp_Skipper-1 HVYTDASDVGSGLMIAQYDDNN---NLRP-------VLFDARKFDSAQR-NYSARDRELLAFIHAVT
Df_Skipper-1 VIHTDASNYGVGSVVSQ----------LDEDGEYHPVIFESRKMTPAER-NYSTTDRELLAAIHVLQ
Dd_Skipper-2 -VYIDASDIGTGCVIMQDDGNG---NIRP-------ILYDSCRFNKFLR-NYSTTDREFLALINVLD
Dp_Skipper-2 HVYIDASDVGTGCMITQFDSDN---NERP-------VLYDSKKFNSFQK-SYTTTDRELLALINVLK
Pp_Skipper-2 FVYTDASDIGISVVICQKDEKG---VLKP-------VLFSGRRLKPGEQ-IKSIGARELLAIKYFLK
Df_Skipper-2 TVTTDASDFGIGAVLEQ-----------DGR----PVAFESRKMTKAER-GMSTRDKEALAVYHSFK
Pf_Skipper-1.1 WLMTDASKVGIGAVLLQGEDWRT----------ARPCGFWSRQYISAER-NYPTHEQELLGVVAAMK
Pf_Skipper-1.2 FLVTDASQFGCGGWLGQGANLES----------ARPARFFSTKFNRAQH-VYTTTDQELLGVLNGCR
Pf_Skipper-1.3 YLFTDASIFGCGGWLGQGPDAES----------VRPFRYFSAKFNSAQR-NYTTTDQELLGVLNGCK
Dd_DGLT-A VIETDASIIGIGGVMIQ-NGK--------------PVSYYSRTLNNAEK-NYSVTERECLAIIESIK
Dp_DGLT-A1 IIETDASNIGIGAVLIQ-NNN--------------IVAVFSEALNDAQT-RYTITEKECLAIVQAIK
Dp_DGLT-A2 TIETDASGIGIAAVLLQ-DGK--------------PVLNFSQLLPPRVK-EIPTIEREAKAVIASLI
Dp_DGLT-A3 VIECDASGIGLAGVLLQ-ENK--------------VVKFYSRTLNSAER-NYSTTEREALAMVECHK
Dl_DGLT-A1 ELHTDASMKGMGAVLYQ-DKK--------------IVAFFSKTYSSSES-NWSASEFECYAVILALK
Dl_DGLT-A2 VLETDASGVGVGATLLQ-DSK--------------VISWFSRKFNENEKKILTACEKECYAIVLALK
Pp_DGLT-A1 HIDCDASNDGIGHVIYQY--KD---NIEQEDNKQIVLYGSKKFNTTER-DYHVFEQEVMAIKHALE
Pp_DGLT-A2 HLECDASKYGIGHALYQINKLD---NITR----DFISFGSRKLTISEI-NYTVLEKELLSIIHALK
Pp_DGLT-A3 IFECDASDVGTGYRLFQL--VD---GKE-----NVICYGSRKLKDSET-RYSTLEKELLAIVTALK
Pf_DGLT-A IIETDASVEGFGAVLAQRFSYPHSETGKLVKEVHPVEYISRSTKPSEK-RYSAFLLELVAVKWALE
Sc_Ty3 RLTTDASKDGIGAVLEEVDNKN---KLVG-----VVGYFSKSLESAQK-NYPAGELELLGIIKALH
Dd_Skipper-1 KFQHLLIGKKVSIYTDHQNLTYIINKSNDKPFTKRQDNYMKYIKEFDYELRHISGKKNGIADFLSR
Dp_Skipper-1 KFQYLLLGKKVTIYSDHQNLTYIINKSNDKPFTKRQDGYMKYIKEFEYELKHISGKQNGIADFLSR
Dl_Skipper-1 KWRRYIHGKEIEIFTDHQPLTRI--LEQPMIQTTHQARRIEFLSEFNIKFKHIDGKANVVADALSR
Pp_Skipper-1 RYGYLLS-RPFVFHTDHKNLIYNSQND---MDNPRLVRWSEILSRFSFQTSYIPGKENCMADYLSR
Df_Skipper-1 KYRHHFLGRKVTIYTDHNNLTYW--DSKHQLS-ARERNWLEILGEFDIKFEYIPGETNTVADALSR
Dd_Skipper-2 RHGYMLKDKKFRLFTDHLNLTYYEELK---EPTKRIIRYLDKASEYKFEVIHVRGEDNNIADMLSR
Dp_Skipper-2 KFDYLLLGNKFIVYSDHLNLAHYQTKQ---ETPKRLIRWLDLISMFSYELVHIEGKKNIAADFLSR
Pp_Skipper-2 KHSDILLGPKLNIYTDHFNLLYLKDQA---ILTKDQLKALEVLSQFNYEINHIEGKNNVGADVLSR
Df_Skipper-2 KWRHYLVGKDFILFSDHTAISR---IEDAVGTDDKVGRWARFIGEFRYTFKHKPGKLNVVADALSR
Pf_Skipper-1.1 AWRLELLGVKFKVLTDHDTLRHFKTQQT---LSKRQARWMETLADYDYELAYIPGKKNAVADAMSR
Pf_Skipper-1.2 KMHEHLIGSHFTVVCDHEPLKTYWSQPPK--QDRRRERLWETLAEYDFDWLFTPGKKNVIADSLSR
Pf_Skipper-1.3 KMHEHLVGWHFVVRCDHEPLKTYWEQPPK--QTRRSTRTWEQLAEYDFEWSFIPGKVNMAADVLSR
Dd_DGLT-A YFKSYINGRKVKIITDHQPLKYLLTGKF---TDR-ITRWTVLLQEYNYEIIYRPGKENFLADALSR
Dp_DGLT-A1 KFRGYINGRKIRIITDHQPLVYIKNGKF---TSR-ISKWSIYLSEYDYEIIYRPGKVNYFADLLSR
Dp_DGLT-A2 YFHPYIYGKEVTIFTDHKPLEAANKIPQ---TNRTLIKLSYILTDYRAQIVYKPGKLNYLADAMSR
Dp_DGLT-A3 AFSSYIKGRKVTIITDHAPLVHLSTGSF---SPR-VQRITLYLQEQDVEIRYRPGEQNILADTLSR
Dl_DGLT-A1 HFHHLIACTSITIESDNLALESLKTGNI---KNKRLAKWSMEIAAYDYHIKHREGRKNVVPDVLSR
Dl_DGLT-A2 HFHYIVASSPITVITDNNALLALKDGKI---KNSRLLKWQLTLAEYQITLVHKPGKLNVVPDALSR
Pp_DGLT-A1 SNYHMLLGYKIVIHTDHQNILFINNKLN-DNTKPKLIRWLQYIFSFNPTLIYKKGSDNVIADGLSR
Pp_DGLT-A2 VNYYHLIGHEVIINTDHKNIKFLREQYA-IGINSRINRWLQFIELFNPTLQYIKGETNVIADGLSR
Pp_DGLT-A3 ANYYHIFGRDIEIRTDHKNITYINEANK-IGINQTINRWIAHINLYQPKIKFIKGKDNTIADGLSR
Pf_DGLT-A KFKPYVFGRPIELISDCQALAGI--LSLQSVSPAHSRWREYILGHDIIKFTHRPGKANAAADGLSR
Sc_Ty3 HFRYMLHGKHFTLRTDHISLLSLQNKNE---PARRVQRWLDDLATYDFTLEYLAGPKNVVADAISR

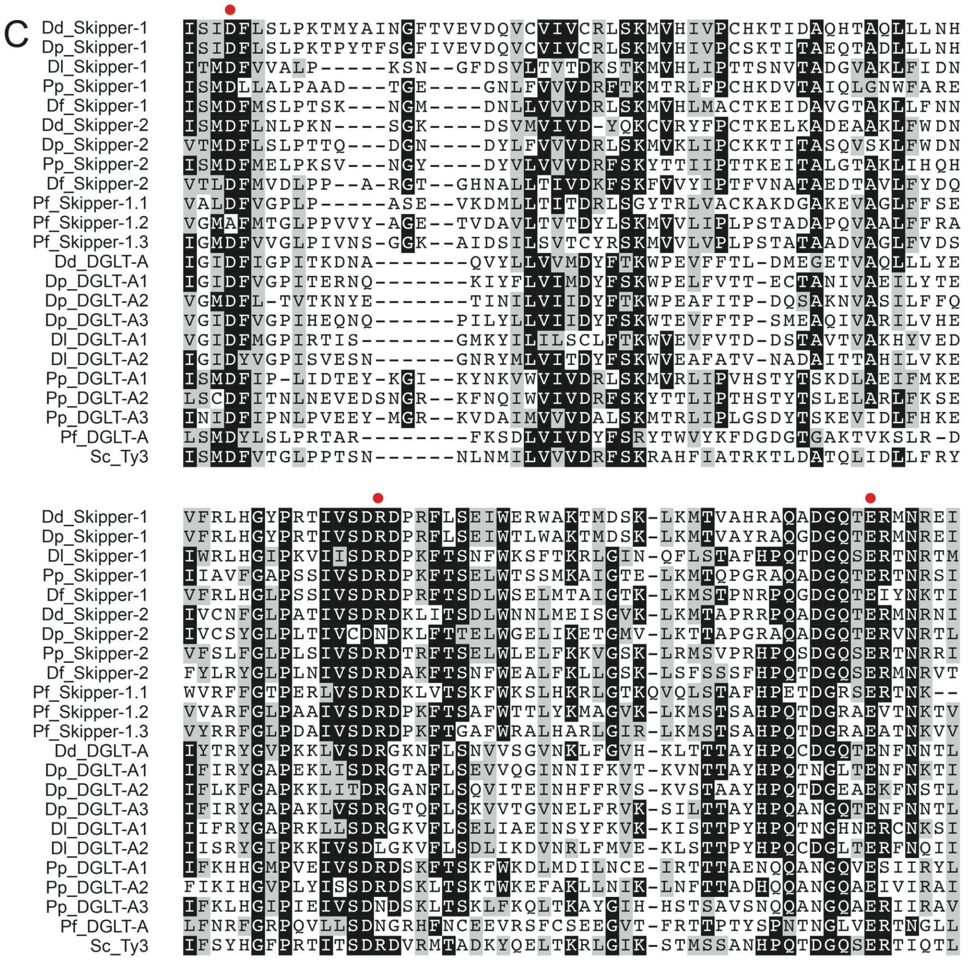

C
Dd_Skipper-1 ISIDFLSLPKTMYAINGFTVEVDQVCVIVCRLSKMVHIVPCHKTIDAQHTAQLLLNH
Dp_Skipper-1 ISIDFLSLPKTPYTFSGFIVEVDQVCVIVCRLSKMVHIVPCSKTITAEQTADLLLNH
Dl_Skipper-1 ITMDFVVALP-----KSN--GFDSVLTVTDKSTKMVHLIPTTSNVTADGVAKLFIDN
Pp_Skipper-1 ISMDLLALPAAD---TGE----GNLFVVVDRFTKMTRLFPCHKDVTAIQLGNWFARE
Df_Skipper-1 ISMDFMSLPTSK---NGM----DNLLVVVDRLSKMVHLMACTKEIDAVGTAKLLFNN
Dd_Skipper-2 ISMDFLNLPKN----SGK----DSVMVIVD-YQKCVRYFPCTKELKADEAAKLFWDN
Dp_Skipper-2 VTMDFLSLPTTQ---DGN----DYLFVVVDRLSKMVKLIPCKKTITASQVSKLFWDN
Pp_Skipper-2 ISMDFMELPKSV---NGY----DYVLVVVDRFSKYTTIIPTTKEITALGTAKLIHQH
Df_Skipper-2 VTLDFMVDLPP--A-RGT--GHNALLTIVDKFSKFVVYIPTFVNATAEDTAVLFYDQ
Pf_Skipper-1.1 VALDFVGPLP-----ASE--VKDMLLTITDRLSGYTRLVACKAKDGAKEVAGLFFSE
Pf_Skipper-1.2 VGMAFMTGLPPVVY-AGE--TVDAVLTVTDYLSKMVVLIPLPSTADAPQVAALFFRA
Pf_Skipper-1.3 IGMDFVVGLPIVNS-GGK--AIDSILSVTCYRSKMVVLVPLPSTATAADVAGLFVDS
Dd_DGLT-A IGIDFIGPITKDNA--------QVYLLVVMDYFTKWPEVFFTL-DMEGETVAQLLLYE
Dp_DGLT-A1 IGIDFVGPITERNQ--------KIYFLVIMDYFSKWPELFVTT-ECTANIVAEILYTE
Dp_DGLT-A2 VGMDFL-TVTKNYE--------TINILVIIDYFTKWPEAFITP-DQSAKNVASILFFQ
Dp_DGLT-A3 VGIDFVGPIHEQNQ-------PILYLLVIIDYFSKWTEVFFTP-SMEAQIVARILVHE
Dl_DGLT-A1 VGIDFMGPIRTIS--------GMKYILILSCLFTKWVEVFVTD-DSTAVTVAKHYVED
Dl_DGLT-A2 IGIDYVGPISVESN-------GNRYMLVITDYFSKWVEAFATV-NADAITTAHILVKE
Pp_DGLT-A1 ISMDFIP-LIDTEY-KGI--KYNKVWVIVDRLSKMVRLIPVHSTYTSKDLAEIFMKE
Pp_DGLT-A2 LSCDFITNLNEVEDSNGR--KFNQIWVIVDRFSKYTTLIPTHSTYTSLELARLFKSE
Pp_DGLT-A3 INIDFIPNLPVEEY-MGR--KVDAIMVVVDALSKMTRLIPLGSDYTSKEVIDLFHKE
Pf_DGLT-A LSMDYLSLPRTAR--------FKSDLVIVDYFSRYTWVYKFDGDGTGAKTVKSLR-D
Sc_Ty3 ISMDFVTGLPPTSN-------NLNMILVVVDRFSKRAHFIATRKTLDATQLIDLLFRY
Dd_Skipper-1 VFRLHGYPRTIVSDRDPRFLSEIWERWAKTMDSK-LKMTVAHRAQADGQTERMNREI
Dp_Skipper-1 VFRLHGYPRTIVSDRDPRFLSEIWTLWAKTMDSK-LKMTVAYRAQGDGQTERMNREI
Dl_Skipper-1 IWRLHGIPKVIISDRDPKFTSNFWKSFTKRLGIN-QFLSTAFHPQTDGQSERTNRTM
Pp_Skipper-1 IIAVFGAPSSIVSDRDPKFTSELWTSSMKAIGTE-LKMTQPGRAQADGQTERTNRSI
Df_Skipper-1 VFRLHGLPSSIVSDRDPRFTSDLWSELMTAIGTK-LKMSTPNRPQGDGQTEIYNKTI
Dd_Skipper-2 IVCNFGLPATIVSDRDKLITSDLWNNLMEISGVK-LKMTAPRRPQADGQTERMNRNI
Dp_Skipper-2 IVCSYGLPLTIVCDNDKLFTTELWGELIKETGMV-LKTTAPGRAQADGQTERVNRTL
Pp_Skipper-2 VFSLFGLPLSIVSDRDTRFTSELWLELFKKVGSK-LRMSVPRHPQSDGQSERTNRRI
Df_Skipper-2 FYLRYGLPLNIVSDRDAKFTSNFWEALFKLLGSK-LSFSSSFHPQTDGQSERMNRVT
Pf_Skipper-1.1 WVRFFGTPERLVSDRDKLVTSKFWKSLHKRLGTKQVQLSTAFHPETDGRSERTNK--
Pf_Skipper-1.2 VVARFGLPAAIVSDRDPKFTSAFWTTLYKMAGVK-LKMSTSAHPQTDGRAEVTNKTV
Pf_Skipper-1.3 VYRRFGLPDAIVSDRDPKFTGAFWRALHARLGIR-LKMSTSAHPQTDGRAEATNKVV
Dd_DGLT-A IYTRYGVPKKLVSDRGKNFLSNVVSGVNKLFGVH-KLTTTAYHPQCDGQTENFNNTL
Dp_DGLT-A1 IFIRYGAPEKLISDRGTAFLSEVVQGINNIFKVT-KVNTTAYHPQTNGLTENFNKTI
Dp_DGLT-A2 IFLKFGAPKKLITDRGANFLSQVITEINHFFRVS-KVSTAAYHPQTDGEAEKFNSTL
Dp_DGLT-A3 IFIRYGAPAKLVSDRGTQFLSKVVTGVNELFRVK-SILTTAYHPQANGQTENFNNTL
Dl_DGLT-A1 IIFRYGAPRKLLSDRGKVFLSELIAEINSYFKVK-KISTTPYHPQTNGHNERCNKSI
Dl_DGLT-A2 IISRYGIPKKIVSDLGKVFLSDLIKDVNRLFMVE-KLSTTPYHPQCDGLTERFNQII
Pp_DGLT-A1 IFKHHGMPVEIVSDRDSKFTSKFWKDLMDILNCE-IRTTTAENQQANGQVESIIRYL
Pp_DGLT-A2 FIKIHGVPLYISSDRDSKLTSKTWKEFAKLLNIK-LNFTTADHQQANGQAEIVIRAI
Pp_DGLT-A3 IFKLHGIPIEIVSDNDSKLTSKLFKQLTKAYGIH-HSTSAVSNQQANGQAERIIRAV
Pf_DGLT-A LFNRFGRPQVLLSDNGRHFNCEEVRSFCSEEGVT-FRTTPTYSPNTNGLVERTNGLL
Sc_Ty3 IFSYHGFPRTITSDRDVRMTADKYQELTKRLGIK-STMSSANHPQTDGQSERTIQTL

Figure S3: Phylogenetic analysis of dictyostelid LTR elements.

Alignment of the concatenated core domains of RT, RNH, and IN was generated with CLUSTALX and analyzed using the Neighbor Joining method. Numbers next to each node indicate bootstrap values as percentages out of 1000 replicates. All positions containing gaps and missing data were eliminated. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. Retrotransposons integrating upstream and downstream of tRNA genes are indicated in red and blue boxes, respectively. Dd: *D. discoideum*; Dp: *D. purpureum*; Dl: *D. lacteum*; Df: *D. fasciculatum*; Pp: *P. pallidum*; Pf: *Protostelium fungivorum*.

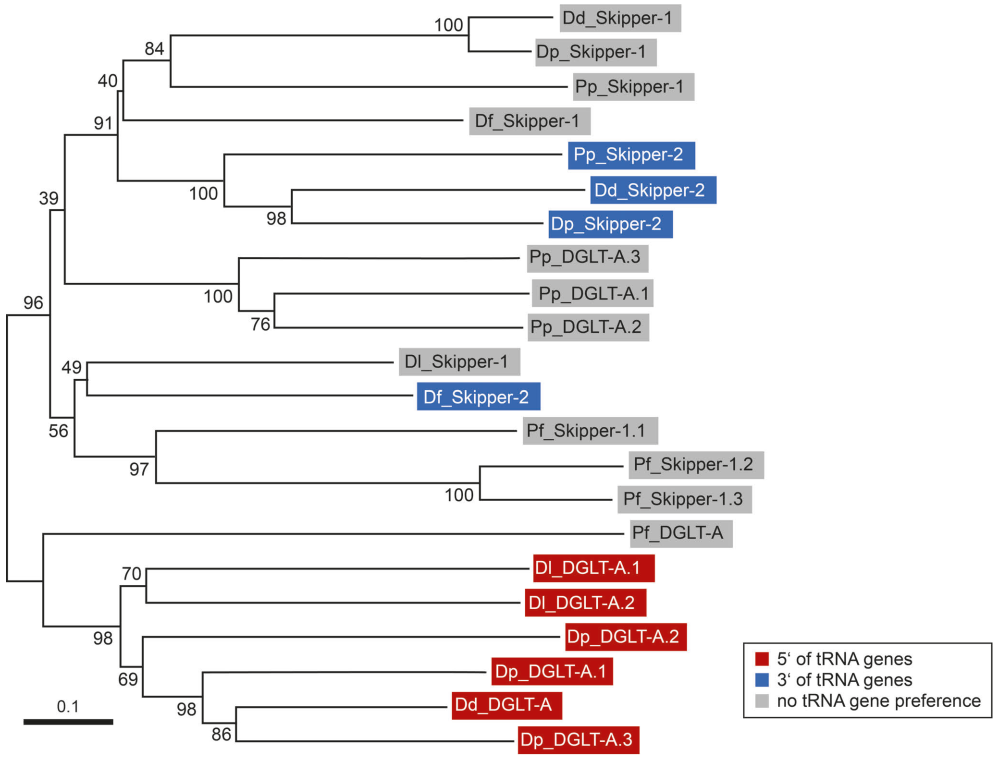
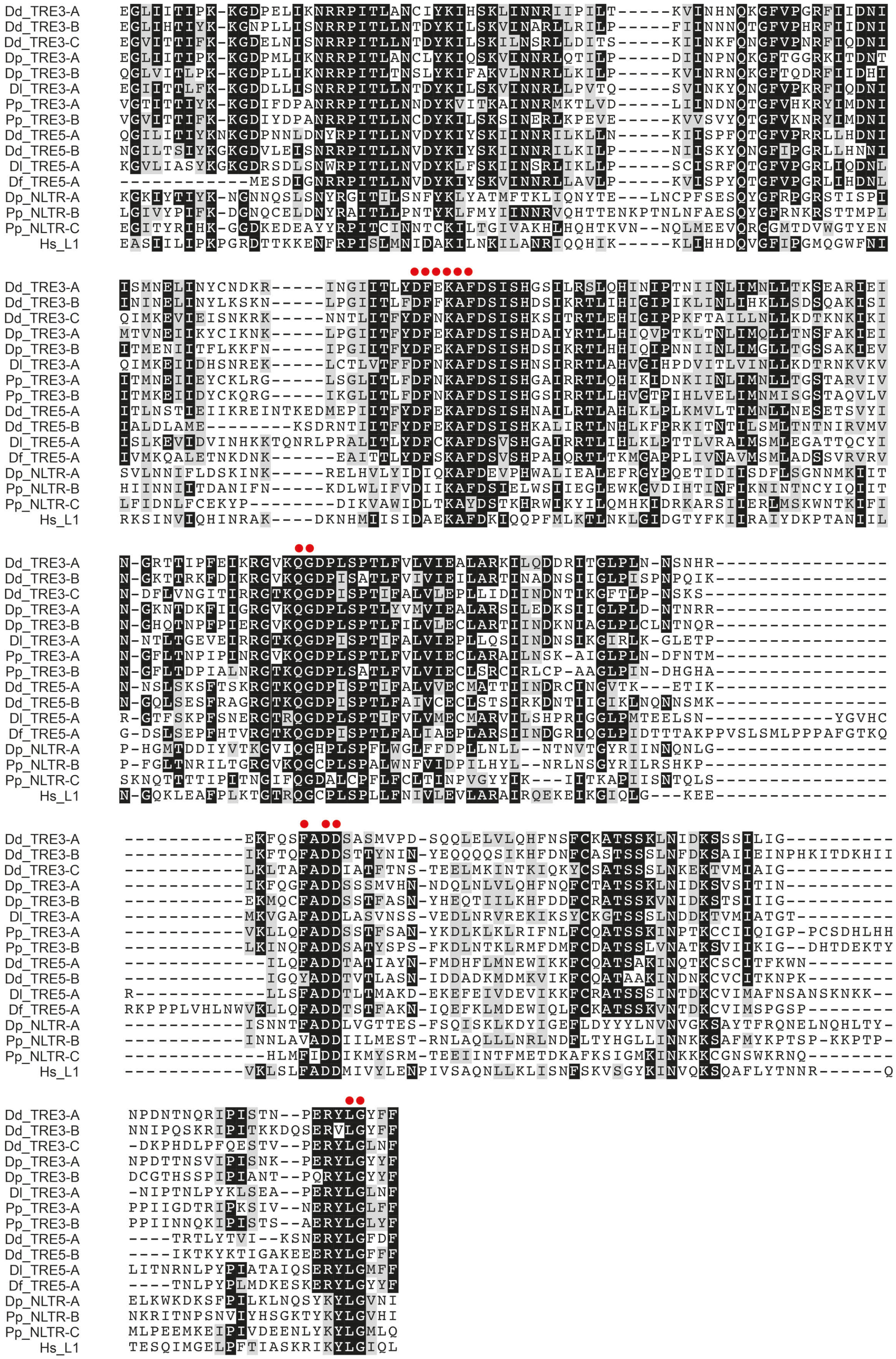


Figure S4: Alignment of RT domains of dictyostelid non-LTR retrotransposons.

Core domains of reverse transcriptases (RT) were determined by searching the Conserved Domain Database [4]. Alignments were generated with CLUSTALX [1]. Shading is to a 50% consensus and was generated with BOXSHADE [2]. Black boxes indicate invariant amino acids and gray boxes represent similar amino acids. Active site residues are indicated by red dots. The sequence of the human L1 element (Hs_L1; U93574) was aligned for comparison. Dd: *Dictyostelium discoideum*; Dl: *D. lacteum*; Dp: *D. purpureum*; Pp: *Polysphondylium pallidum*; Df: *D. fasciculatum.*



```
Dd_TRE3-A   EGLIITIPK-KGDPELIKNRRPITLANCIYKIHSKLINNRIIPILT-----KVINHNQKGFVPGRFIIDNI
Dd_TRE3-B   EGLIHTIYK-KGNPLLISNRRPITLLNTDYKILSKVINARLLRILP-----FIINNNQTGFVPHRFIIDNI
Dd_TRE3-C   EGVITTIFK-KGDELNISNRRPITLLNTDYKILSKILNSRLLDITS-----KIINKFQNGFVPNRFIQDNI
Dp_TRE3-A   EGLIITIPK-KGDPMLIKNRRPITLANCLYKIQSKVINNRLQTILP-----DIINPNQKGFTGGRKITDNT
Dp_TRE3-B   QGLVITLPK-KGDPLLIKNRRPITLTNSLYKIFAKVLNNRLLKILP-----KVINRNQKGFTQDRFIIDHI
Dl_TRE3-A   EGIITTLFK-KGDDLLISNRRPITLLNTDYKILSKVINNRLLPVTQ-----SVINKYQNGFVPKRFIQDNI
Pp_TRE3-A   VGTITTIYK-KGDIFDPANRRPITLLNNDYKVITKAINNRMKTLVD-----LIINDNQTGFVHKRYIMDNI
Pp_TRE3-B   VGVITTIFK-KGDIYDPANRRPITLLNCDYKILSKSINERLKPEVE-----KVVSVYQTGFVKNRYIMDNI
Dd_TRE5-A   QGILITIYKNKGDPNNLDNYRPITLLNVDYKIYSKIINNRILKLLN-----KIISPFQTGFVPRRLLHDNI
Dd_TRE5-B   NGILTSIYKGKGDVLEISNRRPITLLNVDYKIYSKIINNRILKILP-----NIISKYQNGFIPGRLLHNNI
Dl_TRE5-A   KGVLIASYKGKGDRSDLSNWRPITLLNVDYKLFSKIINSRLIKLLP-----SCISRFQTGFVPGRLIQDNL
Df_TRE5-A   ------------MESDIGNRRPITLLNVDYKIYSKVINNRLLAVLP-----KVISPYQTGFVPGRLIHDNL
Dp_NLTR-A   KGKIYTIYK-NGNNQSLSNYRGITILSNFYKLYATMFTKLIQNYTE---LNCPFSESQYGFRPGRSTISPI
Pp_NLTR-B   LGIVYPIFK-DGNQCELDNYRAITLLPNTYKLFMYIINNRVQHTTENKPTNLNFAESQYGFRNKRSTTMPL
Pp_NLTR-C   EGITYRIHK-GGDKEDEAYYRPITCINNTCKILTGIVAKHLHQHTKVN--NQLMEEVQRGGMTDVWGTYEN
Hs_L1       EASIILIPKPGRDTTKKENFRPISLMNIDAKILNKILANRIQQHIK-----KLIHHDQVGFIPGMQGWFNI

Dd_TRE3-A   ISMNELINYCNDKR-----INGIITLYDFEKAFDSISHGSILRSLQHINIPTNIINLIMNLLTKSEARIEI
Dd_TRE3-B   ININELINYLKSKN-----LPGIITLFDFFKAFDSISHDSIKRTLIHIGIPIKLINLIHKLLSDSQAKISI
Dd_TRE3-C   QIMKEVIEISNKRK-----NNTLITFYDFNKAFDSISHKSITRTLEHIGIPPKFTAILLNLLKDTKNKIKI
Dp_TRE3-A   MTVNEIIKYCIKNK-----LPGIITFYDFEKAFDSISHDAIYRTLLHIQVPTKLTNLIMQLLTNSFAKIEI
Dp_TRE3-B   ITMENIITFLKKFN-----IPGIITFYDFEKAFDSISHDSIKRTLHHIQIPNNIINLIMGLLTGSSAKIEV
Dl_TRE3-A   QIMKEIIDHSNREK-----LCTLVTFFDFNKAFDSISHNSIRRTLAHVGIHPDVITLVINLLKDTRNKVKV
Pp_TRE3-A   ITMNEIIEYCKLRG-----LSGLITLFDFNKAFDSISHGAIRRTLQHIKIDNKIINLIMNLLTGSTARVIV
Pp_TRE3-B   ITMKEIIDYCKQRG-----IKGLITLFDFNKAFDSISHGSIRRTLLHVGTPIHLVELIMNMISGSTAQVLV
Dd_TRE5-A   ITLNSTIEIIKREINTKEDMEPIITFYDFEKAFDSISHNAILRTLAHLKLPLKMVLTIMNLLNESETSVYI
Dd_TRE5-B   IALDLAME--------KSDRNTIITFYDFEKAFDSISHKALIRTLNHLKFPRKITNTILSMLTNTNIRVMV
Dl_TRE5-A   ISLKEVIDVINHKKTQNRLPRALITLYDFCKAFDSVSHGAIRRTLIHLKLPTTLVRAIMSMLEGATTQCYI
Df_TRE5-A   IVMKQALETNKDNK------EAITTLYDFSKAFDSVSHPAIQRTLTKMGAPPLIVNAVMSMLADSSVRVRV
Dp_NLTR-A   SVLNNIFLDSKINK-----RELHVLYIDIQKAFDEVPHWALIEALEFRGYPQETIDIISDFLSGNNMKIIT
Pp_NLTR-B   HIINNIITDANIFN-----KDLWLIFVDIIKAFDSIELWSIIEGLEWKGVDIHTINFIKNINTNCYIQIIT
Pp_NLTR-C   LFIDNLFCEKYP--------DIKVAWIDLTKAYDSTKHRWIKYILQMHKIDRKARSIIERLMSKWNTKIFI
Hs_L1       RKSINVIQHINRAK----DKNHMIISIDAEKAFDKIQQPFMLKTLNKLGIDGTYFKIIRAIYDKPTANIIL

Dd_TRE3-A   N-GRTTIPFEIKRGVKQGDPLSPTLFVLVIEALARKILQDDRITGLPLN-NSNHR----------------
Dd_TRE3-B   N-GKTTRKFDIKRGVKQGDPISATLFVIVIEILARTINADNSIIGLPISPNPQIK----------------
Dd_TRE3-C   N-DFLVNGITIRRGTKQGDPISPTIFALVLEPLLIDIINDNTIKGFTLP-NSKS-----------------
Dp_TRE3-A   N-GKNTDKFIIGRGVKQGDPLSPTLYVMVIEALARSILEDKSIIGLPLD-NTNRR----------------
Dp_TRE3-B   N-GHQTNPFPIERGVKQGDPLSPTLFILVLECLARTIINDKNIAGLPLCLNTNQR----------------
Dl_TRE3-A   N-NTLTGEVEIRRGTKQGDPISPTIFALVIEPLLQSIINDNSIKGIRLK-GLETP----------------
Pp_TRE3-A   N-GFLTNPIPINRGVKQGDPLSPTLFVLVIECLARAILNSK-AIGLPLN-DFNTM----------------
Pp_TRE3-B   N-GFLTDPIALNRGTKQGDPLSATLFVLVIECLSRCIRLCP-AAGLPIN-DHGHA----------------
Dd_TRE5-A   N-NSLSKSFTSKRGTKQGDPISPTIFALVVECMATTIINDRCINGVTK---ETIK----------------
Dd_TRE5-B   N-GQLSESFRAGRGTKQGDPISPTLFAIVCECLSTSIRKDNTIIGIKLNQNNSMK----------------
Dl_TRE5-A   R-GTFSKPFSNERGTRQGDPLSPTIFVLVMECMARVILSHPRIGGLPMTEELSN------------YGVHC
Df_TRE5-A   G-DSLSEPFHTVRGTKQGDPISPTLFALIAEPLARSIINDGRIQGLPIDTTTAKPPVSLSMLPPPAFGTKQ
Dp_NLTR-A   P-HGMTDDIYVTKGVIQGHPLSPFLWGLFFDPLLNLL--NTNVTGYRIINNQNLG----------------
Pp_NLTR-B   P-FGLTNRILTGRGVKQGCPLSPALWNFVIDPILHYL--NRLNSGYRILRSHKP-----------------
Pp_NLTR-C   SKNQTTTTIPITNGIFQGDALCPFLFCLTINPVGYYIK---IITKAPIISNTQLS----------------
Hs_L1       N-GQKLEAFPLKTGTRQGCPLSPLLFNIVLEVLARAIRQEKEIKGIQLG---KEE----------------

Dd_TRE3-A   -----------EKFQSFADDSASMVPD-SQQLELVLQHFNSFCKATSSKLNIDKSSSILIG----------
Dd_TRE3-B   -----------IKFTQFADDSTTYNIN-YEQQQQSIKHFDNFCASTSSSLNFDKSAIIEINPHKITDKHII
Dd_TRE3-C   -----------LKLTAFADDIATFTNS-TEELMKINTKIQKYCSATSSSLNKEKTVMIAIG----------
Dp_TRE3-A   -----------IKFQGFADDSSSMVHN-NDQLNLVLQHFNQFCTATSSKLNIDKSVSITIN----------
Dp_TRE3-B   -----------EKMQCFADDSSTFASN-YHEQTIILKHFDDFCRATSSKVNIDKSTSIIIG----------
Dl_TRE3-A   -----------MKVGAFADDLASVNSS-VEDLNRVREKIKSYCKGTSSSLNDDKTVMIATGT---------
Pp_TRE3-A   -----------VKLLQFADDSSTFSAN-YKDLKLKLRIFNLFCQATSSKINPTKCCIIQIGP-PCSDHLHH
Pp_TRE3-B   -----------LKINQFADDSATYSPS-FKDLNTKLRMFDMFCDATSSLVNATKSVIIKIG--DHTDEKTY
Dd_TRE5-A   -------------ILQFADDTATIAYN-FMDHFLMNEWIKKFCQATSAKINQTKCSCITFKWN--------
Dd_TRE5-B   -------------IGQYADDTVTLASN-IDDADKMDMKVIKFCQATAAKINDNKCVCITKNPK--------
Dl_TRE5-A   R------------LLSFADDTLTMAKD-EKEFEIVDEVIKKFCRATSSSINTDKCVIMAFNSANSKNKK--
Df_TRE5-A   RKPPPLVHLNWVKLLQFADDTSTFAKN-IQEFKLMDEWIQLFCKATSSKVNTDKCVIMSPGSP--------
Dp_NLTR-A   -----------ISNNTFADDLVGTTES-FSQISKLKDYIGEFLDYYYLNVNVGKSAYTFRQNELNQHLTY-
Pp_NLTR-B   -----------INNLAVADDIILMEST-RNLAQLLLNRLNDFLTYHGLLINNKKSAFMYKPTSP-KKPTP-
Pp_NLTR-C   -------------HLMFIDDIKMYSRM-TEEIINTFMETDKAFKSIGMKINKKKCGNSWKRNQ--------
Hs_L1       -----------VKLSLFADDMIVYLENPIVSAQNLLKLISNFSKVSGYKINVQKSQAFLYTNNR------Q

Dd_TRE3-A   NPDNTNQRIPISTN--PERYLGYFF
Dd_TRE3-B   NNIPQSKRIPITKKDQSERVLGYFF
Dd_TRE3-C   -DKPHDLPFQESTV--PERYLGLNF
Dp_TRE3-A   NPDTTNSVIPISNK--PERYLGYYF
Dp_TRE3-B   DCGTHSSPIPIANT--PQRYLGYYF
Dl_TRE3-A   -NIPTNLPYKLSEA--PERYLGLNF
Pp_TRE3-A   PPIIGDTRIPKSIV--NERYLGLFF
Pp_TRE3-B   PPIINNQKIPISTS--AERYLGLYF
Dd_TRE5-A   ----TRTLYTVI--KSNERYLGFDF
Dd_TRE5-B   ----IKTKYKTIGAKEEERYLGFFF
Dl_TRE5-A   LITNRNLPYPIATAIQSERYLGMFF
Df_TRE5-A   ----TNLPYPLMDKESKERYLGYYF
Dp_NLTR-A   ELKWKDKSFPILKLNQSYKYLGVNI
Pp_NLTR-B   NKRITNPSNVIYHSGKTYKYLGVHI
Pp_NLTR-C   MLPEEMKEIPIVDEENLYKYLGMLQ
Hs_L1       TESQIMGELPFTIASKRIKYLGIQL
```

Figure S5: Phylogenetic analysis of RT domains of dictyostelid non-LTR elements.

Alignment of the RT domains was generated and analyzed with MEGA7 [5] using the Neighbor Joining method. All positions containing gaps and missing data were eliminated. Numbers next to each node indicate bootstrap values as percentages out of 1000 replicates [6]. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The tree was rooted on RT domains of group II introns.

Sequences used for alignment with dictyostelid NLTR elements were chosen according to a previous phylogenetic analysis performed by Malik et al. [7]: *Trypanosoma brucei* SLACS (X17078), *Trypanosoma cruzi* CZAR (M62862), *Crithidia fasciculata* CRE1 (M33009), *Crithidia fasciculata* CRE2 (U19151), *Drosophila melanogaster* R2 (X51967), *Drosophila mercatorum* R2 (AF015685), *Forficula auricularia* R2 (AF015819), *Bombyx mori* R2 (M16558), *Porcellio scaber* R2 (AF015818), *Bombyx mori* Dong (L08889), *Ascaris lumbricoides* R4 (U29445), *Rattus norvegicus* L1 (U83119), *Mus musculus* L1 (AF081114), *Canis lupus* L1 (AB012223), *Homo sapiens* L1 (U93574), *Oryzias latipes* swimmer (AF055640), *Cyprinodon macularius* swimmer (AF055643), *Xenopus laevis* Tx1 (M26915), *Zea mays* Cin4 (Y00086), *Arabidopsis thaliana* Ta11 (AAC13599), *Dictyostelium discoideum* TRE5-A (X57034), *Chlorella vulgaris* Zepp (AB008896), *Aedes aegypti* JAM1 (Z86117), *Caenorhabditis elegans* RTE1 (AF025462), *Colletotrichum gloeosporioides* Cgt1-3 (L76169), *Magnaporthe grisea* Mgr583 (AF018033), *Ascobolus immersus* Mars1 (X99080), *Neurospora crassa* Tad1 (L25662), *Drosophila melanogaster* R1 (X51968), *Bradysia coprophila* R1 (L00945, *Bombyx mori* R1 (M19755), *Bombyx mori* TRAS1 (D38414), *Aphonopelma sp.* R1 (AF015489), *Bombyx mori* SART1 (D85594), *Anopheles gambiae* RT1 (M93690), *Anopheles gambiae* RT2 (M93691), *Drosophila silvestris* LOA (X60177), *Aedes aegypti* Lian (U87543), *Drosophila miranda* TRIM (X59239), *Drosophila subobscura* Bilbo (U73800), *Drosophila melanogaster* Jockey (M22874), *Drosophila funebris* Jockey (M38437), *Drosophila melanogaster* TART (U14101), *Drosophila melanogaster* Doc (X17551), *Drosophila melanogaster* F (M17214), *Peridroma saucia* YAKPs1, *Drosophila melanogaster* BS (X77571), *Drosophila yakuba* Helena (AF012049), *Drosophila mauritiana* Helena (AF012043), *Bombyx mori* AMY (U07847), *Aedes aegypti* JuanA (M95171), *Culex pipiens* JuanC (M91082), *Chironomus tentans* NCR1Cth (L79944), *Anopheles gambiae* T1 (M93689), *Anopheles gambiae* Q (U03849), *Gallus gallus* CR1 (U88211), *Acanthochelys spixii* CR1 (AB005891), *Schistosoma mansoni* SR1 (U66331), *Biomphalaria glabrata* BGR (X60372), *Drosophila melanogaster* I (M14954), *Drosophila teissieri* I (M28878), *Trypanosoma brucei* ingi (X05710), *Trypanosoma cruzi* L1Tc (X83098). Group II introns: *Streptococcus agalactiae* (WP_000561483), *Leptospira santarosai* (ZP_13300096), *Escherichia coli* (ZP_13884144), *Acetonema longum* (ZP_08623276), *Mycobacterium sp.* (WP_011779439), *Bacillus coagulans* (WP_014095942), *Clostridium kluyveri* (WP_011988638), *Rhodococcus jostii* (WP_011598972), *Frankia sp.* (WP_011436613).

Figure S5 (continued)

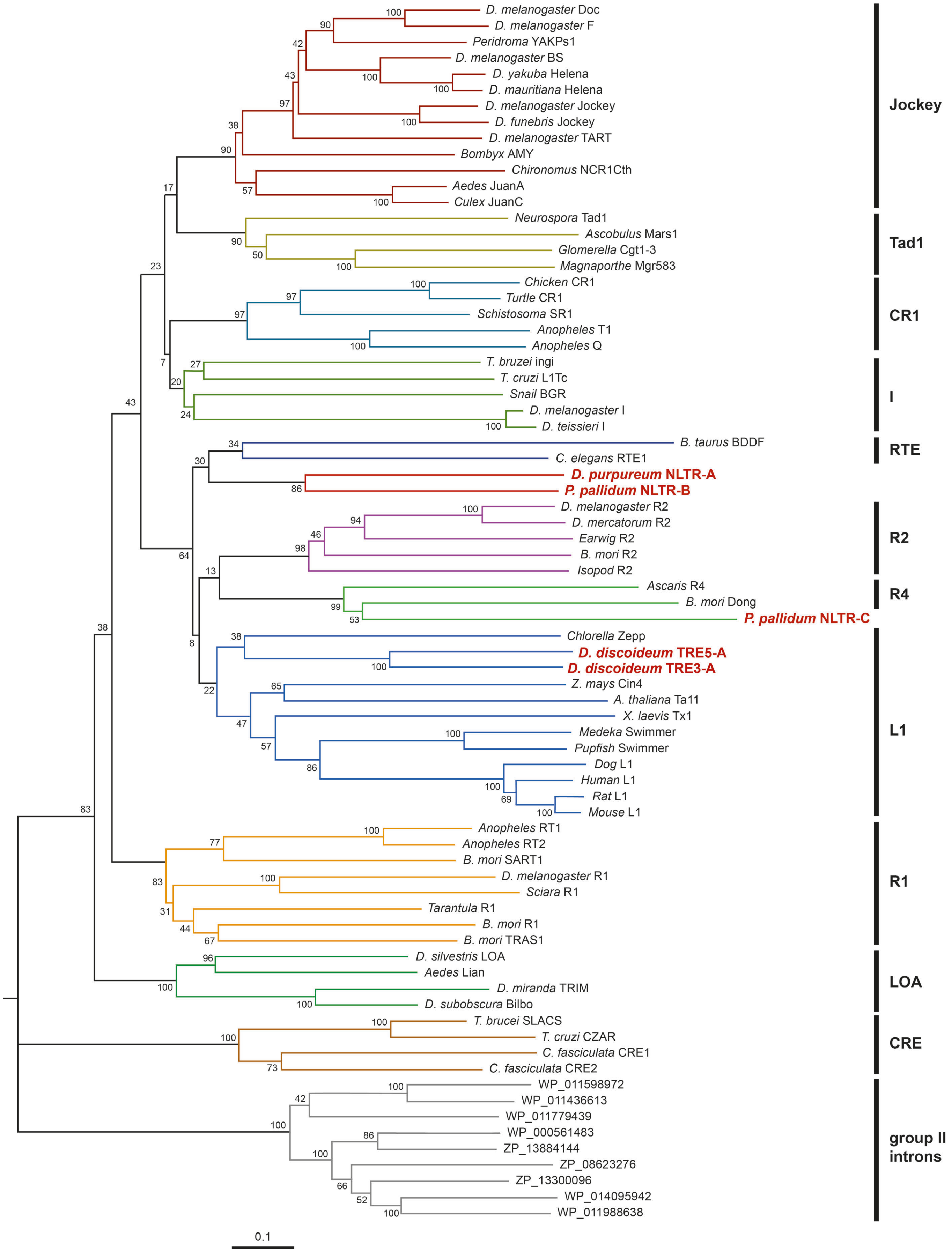
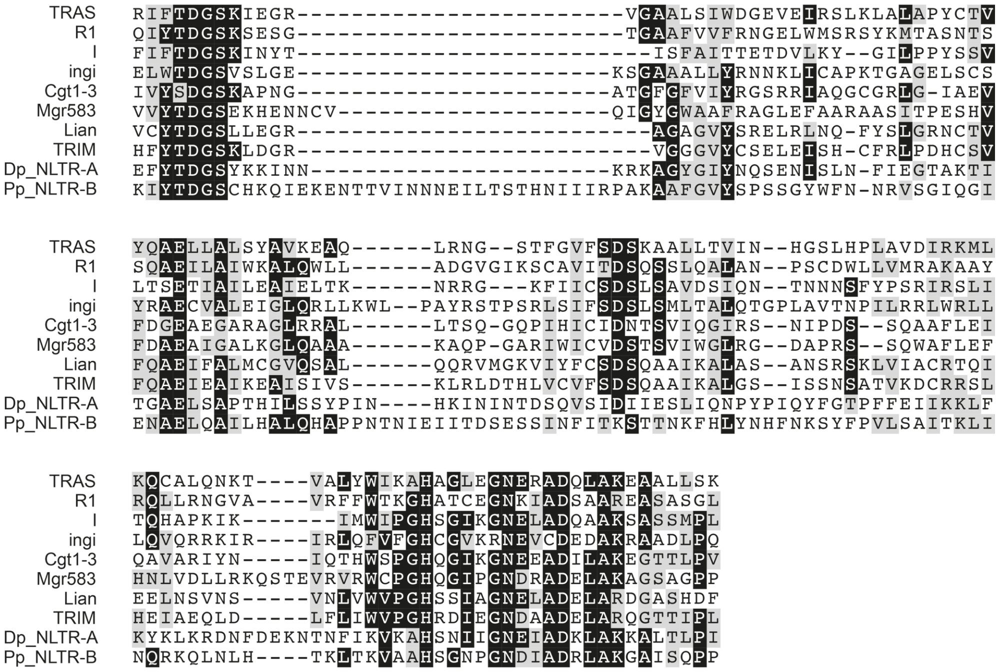


Figure S6: Alignment of dictyostelid non-LTR elements bearing RNH domains.

Core domains of ribonuclease H (RNH) were determined by searching the Conserved Domain Database [4]. Alignments were generated with ClustalX [1]. Shading is to a 50% consensus and was generated with BoxShade [2]. Black boxes indicate invariant amino acids and gray boxes represent similar amino acids. Sequences used for alignment with *D. purpureum* NLTR-A and *P. pallidum* NLTR-B were chosen according to a previous phylogenetic analysis performed by Malik et al. [7]: *Bombyx mori* TRAS (GenBank D38414), *Aphonopelma sp.* R1 (AF015489), *Drosophila melanogaster* I (M14954), *Trypanosoma bruzei* ingi (X05710), *Colletotrichum gloeosporioides* Cgt1-3 (L76169), *Magnaporthe grisea* Mgr583 (AF018033), *Aedes aegypti* Lian (U87543), and *Drosophila miranda* TRIM (X59239).



```
     TRAS  RIFTDGSKIEGR------------------------VGAALSIWDGEVEIRSLKLALAPYCTV
       R1  QIYTDGSKSESG------------------------TGAAFVVFRNGELWMSRSYKMTASNTS
        I  FIFTDGSKINYT--------------------------ISFAITTETDVLKY--GILPPYSSV
     ingi  ELWTDGSVSLGE-----------------------KSGAAALLYRNNKLICAPKTGAGELSCS
   Cgt1-3  IVYSDGSKAPNG-----------------------ATGFGFVIYRGSRRIAQGCGRLG-IAEV
   Mgr583  VVYTDGSEKHENNCV--------------------QIGYGWAAFRAGLEFAARAASITPESHV
     Lian  VCYTDGSLLEGR--------------------------AGAGVYSRELRLNQ-FYSLGRNCTV
     TRIM  HFYTDGSKLDGR--------------------------VGGGVYCSELEISH-CFRLPDHCSV
Dp_NLTR-A  EFYTDGSYKKINN----------------------KRKAGYGIYNQSENISLN-FIEGTAKTI
Pp_NLTR-B  KIYTDGSCHKQIEKENTTVINNNEILTSTHNIIIRPAKAAFGVYSPSSGYWFN-NRVSGIQGI

     TRAS  YQAELLALSYAVKEAQ------LRNG--STFGVFSDSKAALLTVIN--HGSLHPLAVDIRKML
       R1  SQAEILAIWKALQWLL------ADGVGIKSCAVITDSQSSLQALAN--PSCDWLLVMRAKAAY
        I  LTSETIAILEAIELTK------NRRG---KFIICSDSLSAVDSIQN--TNNNSFYPSRIRSLI
     ingi  YRAECVALEIGLQRLLKWL--PAYRSTPSRLSIFSDSLSMLTALQTGPLAVTNPILRRLWRLL
   Cgt1-3  FDGEAEGARAGLRRAL------LTSQ-GQPIHICIDNTSVIQGIRS--NIPDS--SQAAFLEI
   Mgr583  FDAEAIGALKGLQAAA------KAQP-GARIWICVDSTSVIWGLRG--DAPRS--SQWAFLEF
     Lian  FQAEIFALMCGVQSAL------QQRVMGKVIYFCSDSQAAIKALAS--ANSRSKLVIACRTQI
     TRIM  FQAEIEAIKEAISIVS------KLRLDTHLVCVFSDSQAAIKALGS--ISSNSATVKDCRRSL
Dp_NLTR-A  TGAELSAPTHILSSYPIN----HKININTDSQVSIDIIESLIQNPYPIQYFGTPFFEIIKKLF
Pp_NLTR-B  ENAELQAILHALQHAPPNTNIEIITDSESSINFITKSTTNKFHLYNHFNKSYFPVLSAITKLI

     TRAS  KQCALQNKT----VALYWIKAHAGLEGNERADQLAKEAALLSK
       R1  RQLLRNGVA----VRFFWTKGHATCEGNKIADSAAREASASGL
        I  TQHAPKIK-------IMWIPGHSGIKGNELADQAAKSASSMPL
     ingi  LQVQRRKIR----IRLQFVFGHCGVKRNEVCDEDAKRAADLPQ
   Cgt1-3  QAVARIYN-----IQTHWSPGHQGIKGNEEADILAKEGTTLPV
   Mgr583  HNLVDLLRKQSTEVRVRWCPGHQGIPGNDRADELAKAGSAGPP
     Lian  EELNSVNS-----VNLVWVPGHSSIAGNELADELARDGASHDF
     TRIM  HEIAEQLD-----LFLIWVPGHRDIEGNDAADELARQGTTIPL
Dp_NLTR-A  KYKLKRDNFDEKNTNFIKVKAHSNIIGNEIADKLAKKALTLPI
Pp_NLTR-B  NQRKQLNLH----TKLTKVAAHSGNPGNDIADRLAKGAISQPP
```

Table S1: Comparison of retrotransposon in annotated dictyostelid genomes.

	Dictyostelium discoideum	***Dictyostelium purpureum***	***Dictyostelium lacteum***	***Polysphondylium pallidum***	***Dictyostelium fasciculatum***
Haploid genome size (Mb)	34	33	23	33	31
Number of genes	12,646	12,410	10,225	12,675	12,430
Gene density (CDS/Mb)	372	376	444	384	401
Number of tRNA genes	418	353	57	273	198
Retrotransposons total (%)	8.1	0.9	0.000003	0.7	0.6
non-LTR retrotransposons (%)	3.7	0.5	0.000002	0.3	0.3
LTR retrotransposons (%)	1.1	0.4	0.000001	0.3	0.3
YR retrotransposons (%)	3.3	0	0	0	0,1
tRNA gene-targeting retrotransposons (%)	3.8	0.9	0.000002	0.4	0.4
References	[8, 9]	[10]	Genbank LODT01000000	[11]	[11]

CDS: coding sequence; YR: tyrosine recombinase.

Table S2: Association of tRNA genes with TREs in the *D. discoideum* genome.

The individual tRNA gene loci were inspected for the presence of TRE5 or TRE3 elements. The tRNA genes on chromosomes 1-6 were extracted from genome assembly file "*D. discoideum* Non-coding sequences" as of 03-29-2016 (www.dictybase.org). * Note that TRE elements can target the same tRNA genes sequentially, producing tandems. Only the first TRE of such tandems directly adjacent to the targeted tRNA gene was considered in this analysis.

Chromosome	**tRNA genes**	**TREs***	**% tRNA genes occupied***
1	79	52	65.8
2	91	56	61.5
3	79	45	57.0
4	33	21	63.6
5	34	18	52.9
6	87	54	62.1
total	403	246	61.0

Supplemental references

1. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nuc Acids Res. 1997;24:4876-82.
2. BoxShade. http://www.ch.embnet.org/software/BOX_form.html.
3. Malik HS, Eickbush TH. Modular evolution of the integrase domain in the Ty3/gypsy class of LTR retrotransposons. J Virol. 1999;73:5186-90.
4. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015;43(Database issue):D222-D6.
5. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;Mar 22. pii: msw054. [Epub ahead of print].
6. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 1985;39:783-91.
7. Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol. 1999;16:793-805.
8. Glöckner G, Szafranski K, Winckler T, Dingermann T, Quail M, Cox E et al. The complex repeats of *Dictyostelium discoideum*. Genome Res. 2001;11:585-94.
9. Eichinger L, Pachebat JA, Glöckner G, Rajandream M-A, Sucgang R, Berriman M et al. The genome of the social amoeba *Dictyostelium discoideum*. Nature. 2005;435:43-57.
10. Sucgang R, Kuo A, Tian X, Salerno W, Parikh A, Feasley CL et al. Comparative genomics of the social amoebae *Dictyostelium discoideum* and *Dictyostelium purpureum*. Genome Biol. 2011;12:R20.
11. Heidel AJ, Lawal HM, Felder M, Schilde C, Helps NR, Tunggal B et al. Phylogeny-wide analysis of social amoeba genomes highlights ancient originis for complex intercellular communication. Genome Res. 2011;21:1882-91.