

A

>BLAG 1

PPEHRQSILTRQSR TKQSTSSEPQLRNPTDNDQSIVDRQVDVTEINDCSVLNDDLATYKNEMQK
WKLEFSKNNTNDSFNTKVQEFLNFLARAVHLLPGPKHPAVRFYEARRQRRQFSNQVRYKQSSNP
ERGTKRARDKRRELYEYQKTQFLFYNQRRKAVRSALNDRHTACKVDINKIFNYFSSRLSTENSC
ERPAYSSNVTEPERAVLDSTEDSVITKQEVAVAVSKIAADTAPGPDHVLVRALKDATCYEIIAK
IATTMLQHSFVPECLQKARTVL

>BLAG 2

PHIHRSTLVTNLTKNASQPLLDEQADRGSSSNSESSQPDSGFNIALNEWKTKFVRTATANDSDF
DNMVQDFSHFLSTSINLLPGPRHPATKY YEARRKRRLATAEKNYNKSSNP IRASKKAREKRKCK
YDYQLTQYQFYNQRRKAVRRVLDSTAPSKIDLTKVHENFAASF SIENPHIRSNYPSKLT DVER
IELDETYDPTISKEDIAIAAHKIAVDTAPGPDHVLVRVIRNEIVYEILAI IATRMLSSGIVPRC
FKTARTVL

>BLAG 3

GEAYRKKLSHKYECPLENTKDTDTDSPSSNNPITVADVSSPATQQILQFEKALEECILSSDLNR
FEEITVEITAYLSKAIASLPGPKHPATHYYNARKLRQTIHGTSYSNTKNPARLSKRKRKDNKQK
YQYEVAQFHYFNQRRKVVRQIMKNHNNVQCQISIP IIFQHFNTLFNNENPFPTPIESIHPVSTEE
DIEITEDEIKAAMAKIKIDTSAGIDHVLVRTIKELKTSKLLCLLTNAMLKFKICPGILKTARTV
LL

>BLAG 4

PEQHRKSIVMRNTSSVSPVVQ TASKDDNVSNS SANTNSDVGHLHEIRSEVMSFREKFSSLLLS
PCPSKFNEIYNDFTDTLVTINSKLPGPKNPSVRYE LQRNKNIFAANDRKYKDSSNPQKSKR
DRAKRREKFQYELLQYQFYNERKKCVNSLLNDNKRNEPCNIPIGVLKEAMEEKGWTLNTSMLDH
YNNMSPHSQQNDLDQLFTIPVEVIQQSLNKMRTDSAPGPDRI LVRTLKLTECSPVIAVILKIM
LQWGMVPDRMRGARTIL

>BLAG 5

PDLHRLKLVERQTTSSYTIWNSGESSSTCKTSINDTPSKSHPVSVNDPLSNYKQQLLTWTTKFD
CDLTDIDFDTLTNNFLQFLSEAIYYLPGPKHPARKYYALRKKKNKFNCQQSYKNSSNPDRATKR
DRAKRKQKYQYEQTQFDYYNQRRKAVRRVLTSDKPSCQIAVDKVYAHFNSYFSRPNNCIRPDYI
SHVTEVERVELNDSFP THITKEEIKFAISRIAVDTAPGPDHVLMRVLKDDAATGVLSLLATRML
RTGHVPPCLLIARSIL

>BLAG 6

PHIYRKNITERYANASPCLOKTEINNNSGQLKINENKLEPWEKTFSEILNSEKDEVKLDQKLDE
FTLFLCNATSLCKGPVHPATKYKARTTW RPNHEKS NFKTSNNPQRHRR LKEKRQNQYRYDLM
QYYYYYRRKKAVQKIMHLSDYKRC PFSATDVYDYFKETVGTENNNVLHDYSISSPLCNSIDEET
IDYDCQYCTVTEKDIDQAMKGINLDTAPGPDRLVLFKVVKELKCMKIISI IANIMLSWNYVPNSF
QSGRTIL

B

MAFFT-L-INS-i Result

CLUSTAL format alignment by MAFFT (v7.468)

```
Blag_1      PPEHRQSILTRQSRTKQSTSSEPQLRNPTDNDQSIVDRQVDVTEINDCSVLNDDLATYKN
Blag_2      PHIHRSTLVTNLTK-----NASQPLLDEQADRGSSSNSESSQPDSG-----FNI
Blag_5      PDLHRLKLVERQT-----TSSYTIWNSGESSSTCKTSINDTPSKSHPVSVNDPLSNYKQ
Blag_3      GEAYRKKLSHKYE-----CPLENTKDTDTDSPSSNNPIT-----VADVSSPATQ
Blag_6      PHIYRKNITERYA-----NASPCLQKTEINNNSGQLKIN-----EN
Blag_4      PEQHRKSIVMRNTS-----SVSPVVQTASKDDNVSNSANTNSDVGH-----LHEIRS
              :* .: .

Blag_1      EMQKWKLEFSKNNT----NDSFNTKVQEFNLNFLARAVHLLPGPKHPAVRFYEARRQRRQF
Blag_2      ALNEWKTKFVRTAT--ANDSDFDNMVQDFSHFLSTSINLLPGPRHPATKYEARRKRRLA
Blag_5      QLLTWTTKFDCDLT---DIDFDTLTNNFLQLFLSEAIYYLPGPKHPARKYYALRKKNKNF
Blag_3      QILQFEKALEECILS-SDLNRFEEITVEITAYLSKAIASLPGPKHPATHYYNARKLRQTI
Blag_6      KLEPWEKTFSEILNSEKDEVKLDQKLDEFTLFLCNATSLCKGPVHPATKYKARTTWRPN
Blag_4      EVMSFREKFSSLLLSDPCKSKFNEIYNDFDTLVTINSKLPGPKNPSVRYEYELRQRNKNI
              :  :  :          ::  ::  *          ** :*: :*: *  :

Blag_1      --SNQVRYKQSSNPERGTRKARDKRRELYEYQKTQFLFYNQRRKAVRSAL-NDRHTA--C
Blag_2      --TAEKNYNKSSNPIRASKKAREKRCKYDYQLTQYQFYNQRRKAVRRVL-DSTAPS--C
Blag_5      --NCQQSYKNSSNPDRATKRDRAKRKQKYQEQTQFDYYNQRRKAVRRVL-TSDKPS--C
Blag_3      ---HGTSYSNTKNPARLSKRKRKDNKQKYQEVAQFHYFNQRRKVVRQIMKNHNNVQ--C
Blag_6      --HEKSNFKTSNNPQRHRRLKEKRQNQRYDLMQYYYYYRRKKAVQKIMHLSDYKR--C
Blag_4      FAANDRKYKDSSNPQKKSQRDRAKRREKFQYELLQYQFYNERKKCVNSLL-NDNKRNEPC
              :. :.* * :  :: :  .: :  * :  * : :. :.* * .  :          *

Blag_1      KVDINKIFNYFSSRLSTENSCERPAYS-----SNVTEPERAVLDSTEDSVITKQEVAVAV
Blag_2      KIDLTKVHENFAASFSIENPHIRSNYP-----SKLTDVERIELDETYDPTISKEDIAIAA
Blag_5      QIAVDKVYAHFNSYFSRPNNCIRPDYI-----SHVTEVERVELNDSFPTHITKEEIKFAI
Blag_3      QISIPILFQHFNTLFNENPFT-----P-IESIHPVSTEED----IEITEDEIKAAM
Blag_6      PFSATDVYDYFKETVGTENNNVLHDYSISSPLCNSIDEETIDYDCQY-CTVTEKDIDQAM
Blag_4      NIPIGVLKEAMEEKWGTNLNTSMLDHYN-----NSMSPHSQQNDLDQLFTIPVEVIQQSL
              .  :  :  .  *          .          :          :. . :  :

Blag_1      SKIAADTAPGPDHVLVRALKDATCYEIIAKIATTMLQHSFVPECLQKARTVL-
Blag_2      HKIAVDTAPGPDHVLVRVIRNEIVYEILAI IATRMLSSGIVPRCFKTARTVL-
Blag_5      SRIAVDTAPGPDHVLMRVLKDDAATGVL SLLATRMLRTGHVPPCLLIARSIL-
Blag_3      AKIKIDTSAGIDHVLVRTIKELKTSKLLCLLTNAMLKFKICPGILKTARTVLL
Blag_6      KGINLDTAPGPDRVLFKVVKELCKMKIISI IANIMLSWNYVPNSFQSGRTIL-
Blag_4      NKMRTDSAPGPDRILVRTLKLTESCSPVIAVILKIMLQWGMVPDRMRGARTIL-
              :  *::.* *::*.:::          ::. :  .  **          *  :  .*::*
```

C

>Consensus/1-292 Percentage Identity Consensus

PEIHRKKIVERQTKTKQSTNASPCLQDTDADDNSSNNSINDNPDIGHCSSVNDDLSEYKNELLE
WEKKFSECLTSDKDDSDFDEITNEFTAFLSTAISLLPGPKHPATKYEEARKKRRNFFAANQKNY
KNSSNPQRASKRARAKRKEYQYELTQFQFYNQRRKAVRRILNDNSDNPCKIDITKVFEHFNE
KFSTENNCIRPDYSIFSPLSNSTDPERIELDCTYDCTITKEEIAQAMSKIAVDTPGPDHVLVR
VLKDLKCSEIIAIIATIMLQWGIVPDCLKTARTILL

Supplementary Figure 5

Comparative analysis of the amino acid sequences of the BLAG 1 - BLAG 6 retrotransposons located between the zinc finger domains and the reverse transcriptase domain.

A – corresponding amino acid sequences, **B** – MAFFT-L-INS-i alignment, **C** – consensus sequence.