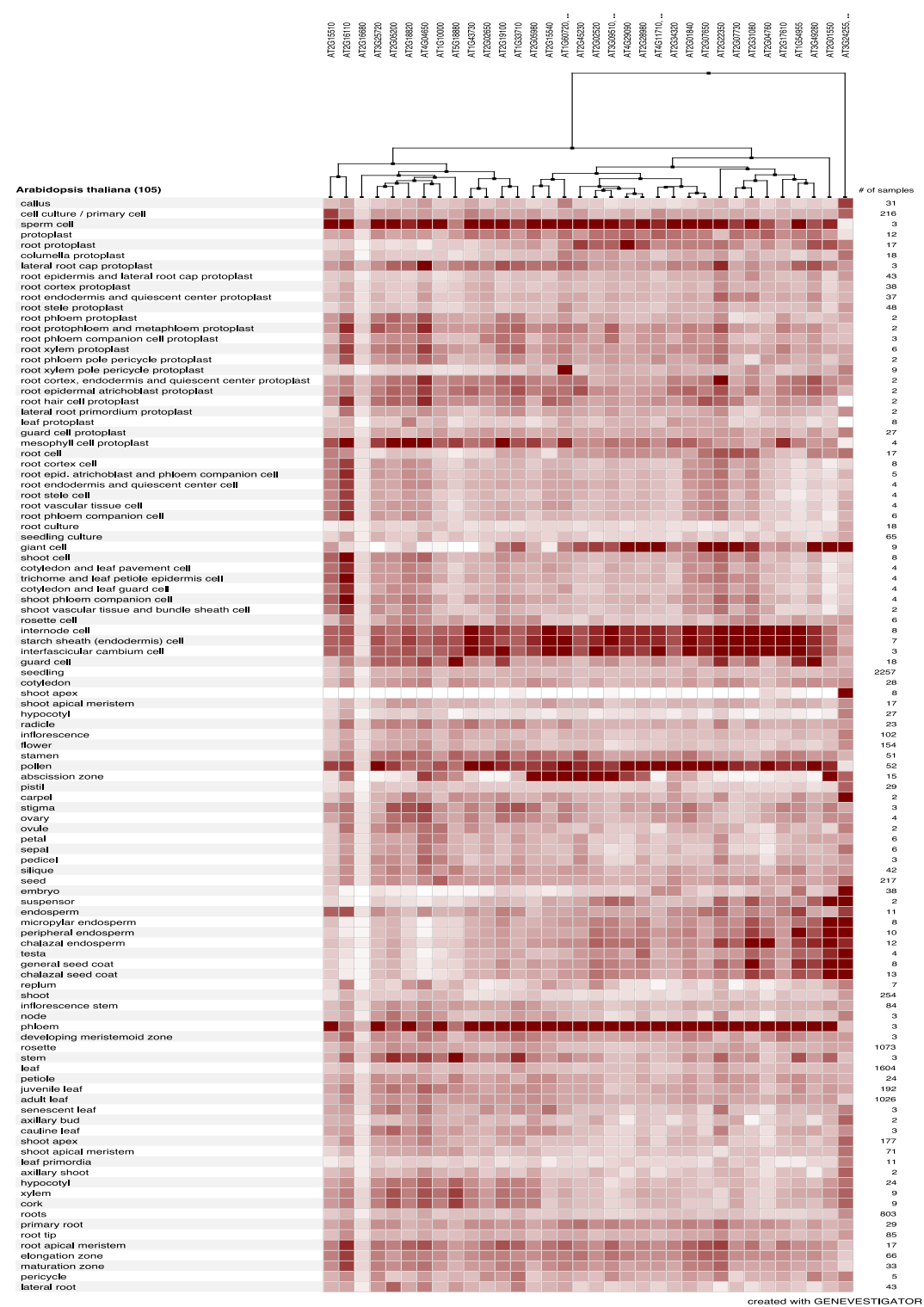


Supplementary materials to Galvan-Gordillo et al., *Biologia* 71, 11, 2016.

Supplementary Table 1. GO annotations of putative endogenized genes from *Arabidopsis*. Expression pattern, subcellular localization and biological roles of the putative endogenized *Arabidopsis* RVT-Znf genes are shown, based on TAIR (<https://www.arabidopsis.org/tools/bulk/go/>) and Genevestigator databases (<https://genevestigator.com/gv/>).

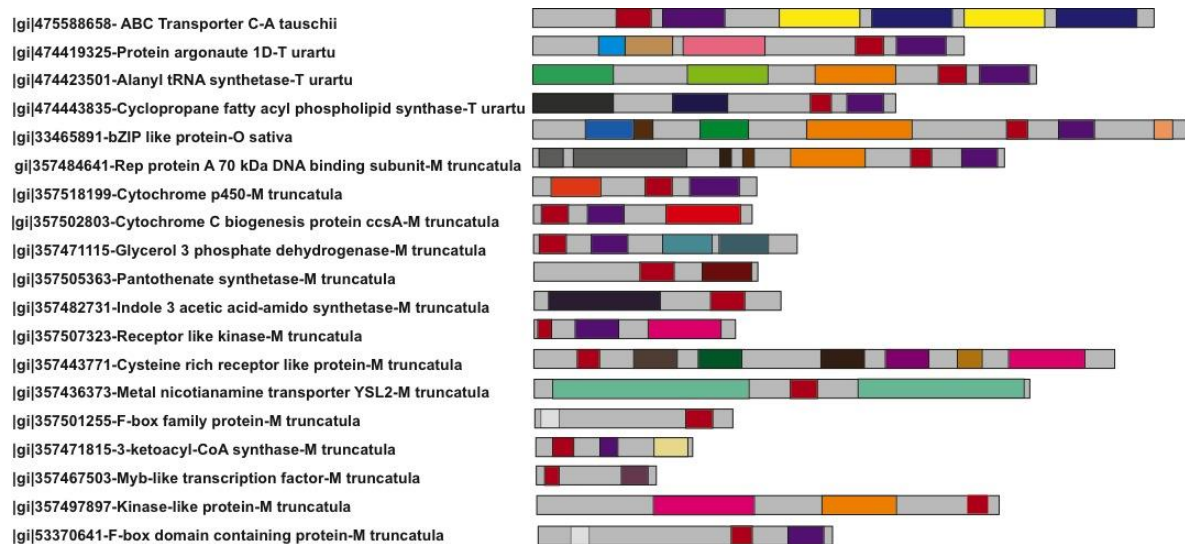
Putative Endogenous Gene	Subcellular Location	Expressed in	Expressed During	Involved in
At1g43730				
At4g04650	Stromule			Response to cold
At2g16110		hypocotyl, petiole, shoot apex		
At2g02520	Mitochondrion	Collective leaf structure, stamen, vascular leaf	LP0.2 two leaves visible stage, petal differentiation and expansion stage	
At1g33710				
At1g60720				
At5g18880	Nucleus			
At3g24255	Nucleus	Guard cell		
At4g11710				
At4g29090	Mitochondrion			
At1g10000		cauline leaf, collective leaf structure, cotyledon, flower, guard cell, hypocotyl, inflorescence meristem, leaf apex, petal, petiole, plant embryo, plant sperm cell, pollen, seed, sepal, shoot apex, shoot system, stem, vascular leaf	LP.02 two leaves visible stage, LP.04 four leaves visible stage, LP.06 six leaves visible stage, LP.10 ten leaves visible stage, LP.12 twelve leaves visible stage, M germinated pollen stage, flowering stage, mature plant embryo stage, petal differentiation and expansion stage, plant embryo bilateral stage, plant embryo cotyledonary stage, vascular leaf senescent stage	
At3g25270	Nucleus			
At3g09510	Mitochondrion			
At2g02650	Mitochondrion	Sperm cell		
At5g36905				
At2g22350				

Supplementary Fig. 1. 105 anatomical parts from 34 Arabidopsis RVT Znf genes were analyzed using the hierarchical clustering tool, using the Pearson correlation, from the GENEVESTIGATOR microarray dataset.

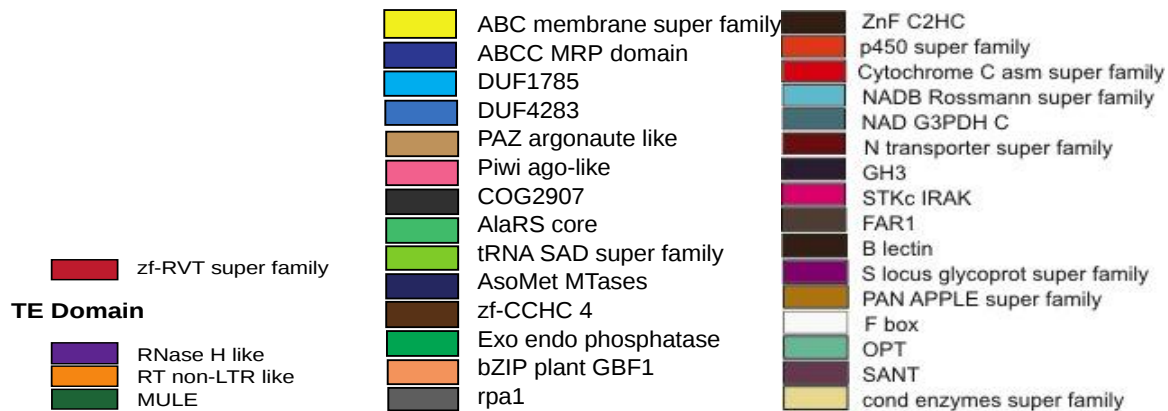


Supplementary Fig. 2. Schematic representation of additional domains in zf-RVTs from plant species.

Genes containing RVT-Znf domain were obtained with the NCBI search and analysis engine using the BLAST tool against the Viridiplantae database. At4g29090 was used as query, and domain structures were obtained using the conserved domain database from NCBI.



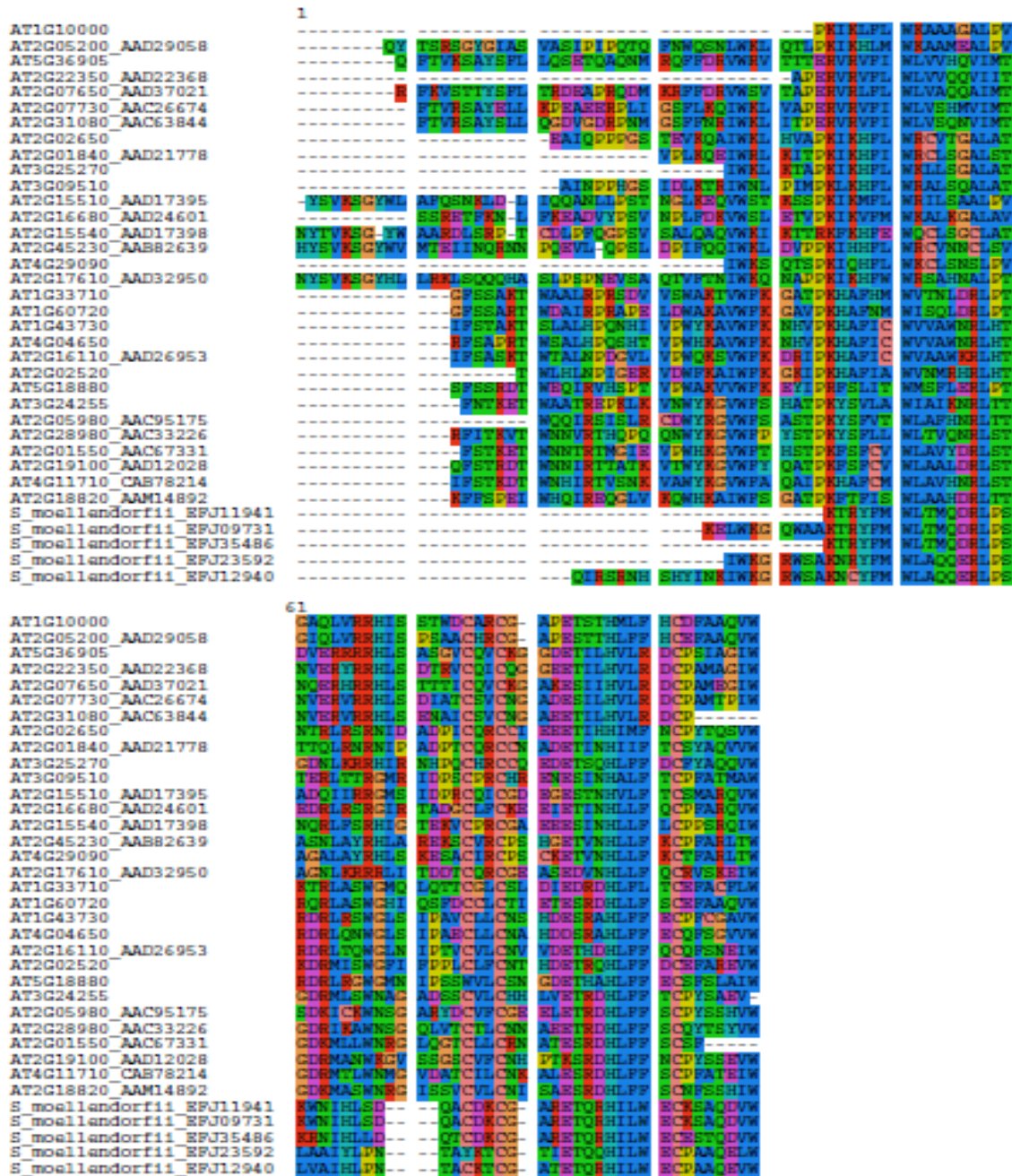
Eukaryotic Domain



Supplementary Fig. 3. Alignment of the RVT-Znf domain containing genes from *Arabidopsis*.

The total of 36 amino acid sequences analyzed were obtained as described in materials and methods, aligned using ClustalX2, manually edited using Seaview.

Alignment: C:\Users\yizuna\Desktop\thesis\ALINEAMIENTOS Y FILOGENIAS DE DOMINIO ZF RVT DE ARABI
Seaview [blocks=10 fontsize=10 A4] on Mon Mar 07 22:10:16 2016

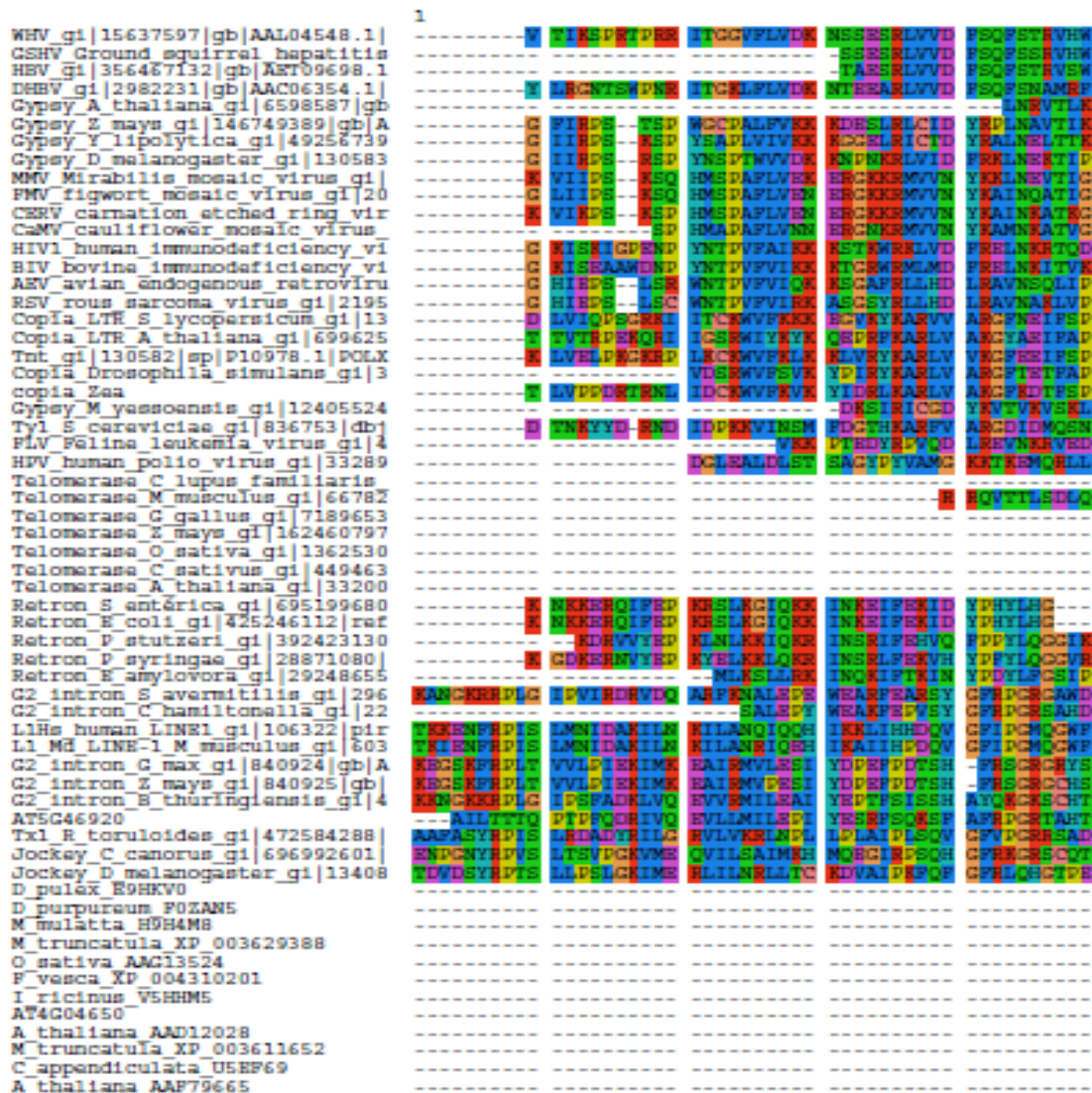


Supplementary Fig 4-8. Alignment of the origin of the RVT-Znf domain.

The total of 60 amino acid sequences analyzed were retrieved from the NCBI and Pfam databases, aligned using ClustalX2 and manually edited using Seaview.

Suppl. Fig. 4.

Alignment: C:\Users\yizuna\Desktop\thesia\filogenias articulo\origen zf-RVT\DominioRT_R_zfRVT_R
Seaview [blocks=10 fontsize=10 A4] on Mon May 23 21:32:33 2016



Suppl. Fig. 5.

101

WHV_gi 15637597 gb AAL04548.1	-----	CS	PCF	IMCFPRILDMC	VCLESPFLLAQ	FTSALASNVH
GSHV Ground squirrel hepatitis	-----	CS	PCF	IMCFPRILDMC	VCLESPFLLAQ	FTSALASNVH
HBV_gi 356467132 gb XET09698.1	-----	CS	SCI	IPCFPRILDMC	VCLESPFLLAQ	FTSALASNVH
DHBV_gi 2982231 gb AAC06354.1	-----	SDC	-	VYFPRILDMC	VCLESPFLLAQ	FTSALASNVH
Cypsey A thaliana gi 6598587 gb	-----	STR	YC	HFEPVVMDFC	LINADTALMH	LMNSVVFQEF
Cypsey X mays gi 146749389 gb A	-----	STR	YC	LYEFLVMDFC	LINADTALMH	LMNSVVFQEF
Cypsey Y lipolytica gi 49256739	-----	STR	SC	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
Cypsey D melanogaster gi 130583	-----	SVN	CC	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
MMV Mirabilis mosaic virus gi	-----	CC	PC	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
PMV figwort mosaic virus gi 20	-----	CC	PC	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
CERV carnation etched ring vir	-----	CC	PC	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
CaMV cauliflower mosaic virus	-----	CC	PC	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
HIV1 human immunodeficiency vi	-----	CI	PS	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
BIV Bovine immunodeficiency vi	-----	SVV	PS	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
AEV avian endogenous retroviru	-----	LVV	PS	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
RSV rous sarcoma virus gi 2195	-----	FLD	PS	HYEFLVMDFC	LINADTALMH	LMNSVVFQEF
Copia LTR S lycopersicum gi 13	-----	CFQ	VH	VCLLNKSLYC	LQAPSPQWYK	FPDSYMWKLC
Copia LTR A thaliana gi 699625	-----	CCS	ES	VCLLNKSLYC	LQAPSPQWYK	FPDSYMWKLC
Tnt gi 130582 sp P10978.1 POLX	-----	CFE	VM	VCLLNKSLYC	LQAPSPQWYK	FPDSYMWKLC
Copia Drosophila simulans gi 3	-----	GIS	CM	VCLLNKSLYC	LQAPSPQWYK	FPDSYMWKLC
copia Zea	-----	CFAD	VI	HCHLNKSLYC	LQAPSPQWYK	FPDSYMWKLC
Cypsey M yassoonensis gi 12405524	-----	NTHE	CC	LFVYNRLPYC	VASADGIFOR	TMSHLLQCF
Ty1 S cerevisiae gi 836753 dbj	-----	HLCD	-	LLRLNKLKSLYC	LQAPSPQWYK	FPDSYMWKLC
PLV Feline leukemia virus gi 4	-----	FWSD	CC	GLTWYRLPYC	FNNSDPLFDS	ALHSPLADP
HPV human polio virus gi 33289	-----	CCD	PS	HHLYKNKPYC	VKSGTSLFNS	MINNLLIRTL
Telomerase C lupus familiaris	-----	HI	CC	CCSYIQCNC	VDCSILSTL	LCSLYFCGLE
Telomerase M musculus gi 66782	-----	HI	CC	CCSYIQCNC	VDCSILSTL	LCSLYFCGLE
Telomerase C gallus gi 7189653	-----	HI	CC	CCSYIQCNC	VDCSILSTL	LCSLYFCGLE
Telomerase X mays gi 162460797	-----	-----	-----	-----	-----	-----
Telomerase O sativa gi 1362530	-----	-----	-----	-----	-----	-----
Telomerase C sativus gi 449463	-----	-----	-----	-----	-----	-----
Telomerase A thaliana gi 33200	-----	-----	-----	-----	-----	-----
Retron S antierica gi 695199680	-----	-----	-----	-----	-----	-----
Retron E coli gi 425246112 ref	DVS	-----	-----	-----	-----	-----
Retron P stutseri gi 392423130	DVS	-----	-----	-----	-----	-----
Retron P syringae gi 28871080	DVS	-----	-----	-----	-----	-----
Retron E amylovora gi 29248655	DVS	-----	-----	-----	-----	-----
G2 intron S avermitilis gi 296	-----	-----	-----	-----	-----	-----
G2 intron C hamiltonella gi 22	-----	-----	-----	-----	-----	-----
L1hs human LINE1 gi 106322 pir	-----	-----	-----	-----	-----	-----
L1 Md LINE-1 M musculus gi 603	-----	-----	-----	-----	-----	-----
G2 intron C max gi 840924 gb A	-----	-----	-----	-----	-----	-----
G2 intron X mays gi 840925 gb	-----	-----	-----	-----	-----	-----
G2 intron B thuringiensis gi 4	-----	-----	-----	-----	-----	-----
AT5G46920	-----	-----	-----	-----	-----	-----
Tx1 R toruloides gi 472584288	-----	-----	-----	-----	-----	-----
Jockey C canorus gi 696992601	-----	-----	-----	-----	-----	-----
Jockey D melanogaster gi 13408	-----	-----	-----	-----	-----	-----
D pulax B9HKV0	-----	-----	-----	-----	-----	-----
D purpureum P0ZAN5	-----	-----	-----	-----	-----	-----
M mulatta H9H4M8	-----	-----	-----	-----	-----	-----
M truncatula XP_003629388	-----	-----	-----	-----	-----	-----
O sativa AAC13524	-----	-----	-----	-----	-----	-----
F vesca XP_004310201	-----	-----	-----	-----	-----	-----
I ricinus V5HHM5	-----	-----	-----	-----	-----	-----
AT4G04650	-----	-----	-----	-----	-----	-----
A thaliana AAD12028	-----	-----	-----	-----	-----	-----
M truncatula XP_003611652	-----	-----	-----	-----	-----	-----
C appendiculata U5EF69	-----	-----	-----	-----	-----	-----
A thaliana AAP79665	-----	-----	-----	-----	-----	-----

Suppl. Fig. 6.

151		
WHV_gi 15637597 gb AAL04548.1	RNPCH	CVVPAYMDDV VLCAKSYEHL TAYYSHICSV PLDLCIRHNV
GSHV Ground squirrel hepatitis	RNPCH	CLAPAYMDDL VLCAKSYEHL TAYYSHICSV PLDLCIRHNV
HBV_gi 356467132 gb AET09698.1	RNPCH	CLAPAYMDDV VLCAKSAQHL ELYTAVTNF LLSDCIRHNV
DHBV_gi 2982231 gb AAC06354.1	RNPCH	VWIFPTNDDF LLCHTNARHL MAISHAVCSF LQELCIINF
Gypsy A thaliana gi 6598587 gb	DEFVH	VFIDDI LVYPSDEEH EVHLRVMEX LRSQALFAL
Gypsy X mays gi 146749389 gb A	DEFVH	VFIDDI LVYKNEREH AEHLRVVLQR LRRHLYAKL
Gypsy Y lipolytica gi 49256739	MTVC	VYLODI IVPSRNRSDH KSHVREVLQA LRAQRFYAKK
Gypsy D melanogaster gi 130583	CHICY	VYVDDV IIPSENERSDH VEHIDVLRK LIDAMREVSQ
MMV Mirabilis mosaic virus gi	KEPSL	VYVDDI IVPSRNRSDH QOHVMVLRK IESLCIILSK
PMV figwort mosaic virus gi 20	DFPCM	VYVDDI IVPSRNRSDH YMHVAVLKI VERTCIILSK
CERV carnation etched ring vir	AKYCC	VYVDDI LVPSDTERDH YIHVILRRE CERLCIILSK
CaMV cauliflower mosaic virus	RFCC	VYVDDI VVPSNNEEDH LLHVAMILKE CMBICIILSK
HIV1 human immunodeficiency vI	LNDE	MVIVQYMDL YVCSOLEICQ HRTXELSLA HLLRWCFTTD
BIV Bovine immunodeficiency vi	LNDD	VMLYQYMDL LIGENDDO HKQIVGRIID NLGSCYCFIV
AEV avian endogenous retrovirus	MTSAG	CIILHYMDL LLAAFLAR LQLECVIS LRAACFTIV
RSV rous sarcoma virus gi 2195	LNDS	ELMLHYMDL LLAASSHDD LEAACEVIE TLERACFTIS
Copia LTR S lycopersicum gi 13	YTESYDCCV	IYVLVYVDDM LIAAKKTYDT KLEKCLSAE FEMKDLCAAE
Copia LTR A thaliana gi 699625	PKRSYDSCA	MYLLPYVDDM LVAANNMQAI DALKKELIK FEMKDLCAAE
Tnt gi 130582 sp P10978.1 POLX	YLYTYDPCV	IILLYVDDM LIVCKDRGLI AALCKDLSES FDMKDLCPAQ
Copia Drosophila simulans gi 3	PVNSVDRCI	IYVLVYVDDV VIAACDMYAM NNPETHLMEK FDMKDLNEIA
Copia Zea	PVDSKADVSL	IYVLVYVDDI IITGSSPHAI ENVLAKLDD PAIKDLCDLH
Gypsy M yassoonensis gi 12405524	QVIVR	VDDI LVTCNKRCH LERLEAVLNR LDRACVLRK
Ty1 Scareviciae gi 836753 dbj	QVQVCHNC	VTTCLFVDDM ILPSKDLMAN KKIITLKAQ YDAIILNCE
FLV Felina leukemia virus gi 4	VRYDA	LVILQYVDDL LLAAATREDC LSCFALLET LCNKCTRASA
HPV human papilloma virus gi 33289	LLATYD	ELMLAYGDDV IASYPHEVLA SLLAQCKDY LTMTPADASA
Telomerase C lupus familiaris	RLFPDCEGD	CVLLRLVDDF LLVTTHLQQA QAFLETLVRC VDETCCANL
Telomerase M musculus gi 66782	NLPAEYORD	GLLLRFVDDF LLVTTHLQQA KTFLETLVRC VDETCCANL
Telomerase C gallus gi 7189653	NLLCCIQKD	CVLLRLVDDF LLVTTHLQQA KTFLETLVRC VDETCCANL
Telomerase X mays gi 162460797	NELLKSLPD	SLIMRFIDDF IPISPSERDA LPPFNMRER PVNTYCYMD
Telomerase O sativa gi 1362530	NSVLSKPLD	YLIMRFIDDF IPISPSERDA KPLNMRER PVNTYCYMD
Telomerase C sativus gi 449463	RVLPDPLPD	YMLRFIDDF LPISSEKQLA KPLCRVHRC FAYTCYMD
Telomerase A thaliana gi 33200	RLITPFLTS	YLLRFIDDF LPSVSRQQA SFPYKLRK FAYTCYMD
Retron S antlerica gi 695199680	YMYNRLMC	LEYSHLLDDI TISSDSSEK KVIKLVHRC VMQYRLSINE
Retron E coli gi 425246112 ref	YMYNRLMC	LEYSHLLDDI TISSDSSEK KVIKLVHRC VMQYRLSINE
Retron P stutzeri gi 392423130	YSLVYFRKC	ITYSHLLDDV TISSDSPEI ERAISVARM FSHHLRIHD
Retron P syringae gi 28871080	YMFVRLKOC	VTYSHLLDDV TISSDSQEV GYIKTVACL FSHHLRIKX
Retron E amylovora gi 29248655	PKVCHLNG	PHYTHLLDDI TISSAKGSM NYVESTICEF ITQKRLSINK
G2 intron S avermitilis gi 296	EVICANRPN	DTIMRYADDF VVFCYENEA IAAQOLAAM LEFSLSPNE
G2 intron C hamiltonella gi 22	LLCIGYWNK	YAVRYADDF VVFCNREBC ETAKIKLOIN LAQCLALSE
L1Hs human LINE1 gi 106322 pir	RAINSEHIX	VVILRFADDM IVYLENPTSA GMLALISNF SEVCHYINV
L1 Md LINE-1 M musculus gi 603	RAIROHIX	VVISLLADDM IVYISDFSV HELNLISNF SEVCHYINV
G2 intron C max gi 840924 gb A	QICIRHORY	----- S L RIRSV LLRTG-SIDD
G2 intron X mays gi 840925 gb	QICIRHOKH	----- S L LVVIRSV LLIRGRIIDD
G2 intron B thuringiensis gi 4	RYMDPIOKH	RYVIALENK MCEASQHLQA KALKELESI PYSNPMDSNY
AT5G46920	RYMETVYDF	DSKSDV IANNDEG AD QNTSNPEFY DTCGPO-ET
Tx1 R toruloides gi 472584288	DALLRQVAL	STTLAFADDA ILACCAIMAL PRLALAQDM ILASCHINT
JocKey C canorus gi 696992601	SCIRCP	DDTLQCSVD LLSCHER-AL QRLNRLDCK ASTNCHPFX
Jockey D melanogaster gi 13408	YHTVYEVDE	YTAVLTKSK ILAATY-CL GEYLCAPOQH AENWYVINA
D pulax E9HKV0	-----	FSVDDF FLPHNRLDY HDRANKIDCT INPECKCKS
D purpureum POZAN5	GWIDN	----- HIRATP WEAVLSLLCN RMELVHRDAS SMADCHLCAE
M mulatta H9B4M8	KVFC	----- L WLLMGNRLNT KPMLEHYNCA PPDCCVLCDA
M truncatula XP 003629388	KNF	----- L WEACHNCLPT HIRLQSKVQ CIDRCVCCD
O sativa AAC13524	KIL	----- L WEAAHECLAT CFQRRHHIF STDCCVFCN
F vesca XP 004310201	KCP	----- I WLLKDSSTT REALASHIAM RYQSCVFCN
I ricinus V5HHM5	KFP	----- F FLKHTNILEV VVMMQCKMF VVMSVFLCA
AT4G04650	PHAP	----- C MVVAMNRLHT RDLQWCLG IPASCILLCA
A thaliana AAD12028	PHPS	----- V WLAALRLST GDRMNRKCV SSCSCVFCN
M truncatula XP 003611652	NIF	----- V WRLPLNRLAY KINLRKRNVL RATNCRALCC
C appendiculata USEP69	NSKK	----- SIL YLLTNERICH KDLYPSQNEI TENLCQCN
A thaliana AAP79665	AS	----- M FLCLVGNQAL MIVVERPWHH RCITCQMKC

Suppl. Fig. 7.

	201				
WHV_gi 15637597 gb AAL04548.1	NTSWACH				
GSHV Ground squirrel hepatitis	EVSWACHTE	HFMCYTI			
HBV_gi 356467132 gb AET09698.1	NTVRRGYSL	NFMCTV			
DHBV_gi 2982231 gb AAC06354.1	DTTTP				
Gypsy A thaliana gi 6598587 gb	GCSPWQREM	DPLCHIV			
Gypsy Z mays gi 146749389 gb A	SCDPLWKEI	KPLCHTI			
Gypsy Y lipolytica gi 49256739	SKCELPHKIM	CPFLCHVV			
Gypsy D melanogaster gi 130583	KRTFFPKESV	EYLCPIV			
MMV Mirabilis mosaic virus gi	KKANLPKKEI	NPLCLEI			
PMV figwort mosaic virus gi 20	KKANLPKKEI	NPLCLEI			
CERV carnation etched ring vir	KKAQLPKKEI	NPL			
CaMV cauliflower mosaic virus	KKAQLPKKEI	NPLCLEI			
HIV1 human immunodeficiency vi	DKKHCKEPFP	LWMCYEL			
BIV bovine immunodeficiency vi	DEIVCEEV				
AEV avian endogenous retroviru	SEIICRSCV	QPLCYAF			
RSV rous sarcoma virus gi 2195	PDIVCKEPCV	QYLCTYL			
Copia LTR S lycopersicum gi 13	NILCEIILID	SEERHLP			
Copia LTR A thaliana gi 699625	NILCEIILID	SEAC			
Tnt gi 130582 sp P10978.1 POLX	QILCMKIVRE	RTSRHLM			
Copia Drosophila simulans gi 3	HPICIRIEMH	ED--HIY			
copia Zaa	YPLCEVHRE	GD--GLL			
Gypsy M yessoensis gi 12405524	SCYFLAKKV	EYLCHRI			
Ty1 S cereviciae gi 836753 dbj	S--DNELQYD	ILCLEIK			
FLV Felina leukemia virus gi 4	KKAQICLOEV	TYLCYSL			
HPV human polio virus gi 33289	TVETIVENV	CPLEHFF			
Telomerase C lupus familiaris	QETAVNFTVE	DCALCSAFL	QLP--AHCLP	PWL	
Telomerase M musculus gi 66782	QETVVFPTVE	PCPLGCA			
Telomerase G gallus gi 7189653	KETVVNFTVD	D--PCCSFF	HLG--DCCLI	SWCCL	
Telomerase Z mays gi 162460797	SKYCFNFEVA	NSEHC-SNRI	YKCDGCFPI	PWSCL	
Telomerase O sativa gi 1362530	SKYCFNFCAG	NSEHC-SNRI	YKCDGCFPI	PWSCL	
Telomerase C sativus gi 449463	SKYCFNFOVA	NTYHIVSKV	HVCHGCVSPL	HWACL	
Telomerase A thaliana gi 33200	TFPCINFED	SEERHCSNEM	FVCHGCVFV	HWACL	
Retron S antartica gi 695199680	SKYTIHESKD	SSSKLSVTCL	WVKHCVPKLT	KENRE	
Retron E coli gi 425246112 ref	SKYTIHESKD	SSSKLSVTCL	WVKHCVPKLT	KENRE	
Retron P stutzeri gi 392423130	KKKIIVSSD	TESYIVTCV	WVKHCVPKLT	KENRE	
Retron P syringae gi 28871080	SKYTIHESKD	LSARYIVTCV	WVKHCVPKLT	KENRE	
Retron E amylovora gi 29248655	DK--VISISS	SLQQFVHSL	SICCHAPFP	KVERD	
G2 intron S avermitilis gi 296	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
G2 intron C hamiltonella gi 22	KETSIHLEK	GFDFLGFN			
L1Hs human LINE1 gi 106322 pir	QSSQAFLYTH	NKQTESQIME	ELPFTIASR	LYLCL	
L1 Md LINE-1 M musculus gi 603	NKQMAFLYTH	NKQAKKEIR	TFPFSIVTNN	LYLCL	
G2 intron C max gi 840924 gb A	QENSCEASL	NAPD			
G2 intron Z mays gi 840925 gb	QETVCEASL	NAPD			
G2 intron B thuringiensis gi 4	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
AT5G46920	KETSIHLEK	GFDFLGFN			
Tx1 R toruloides gi 472584288	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
Jockey C canorus gi 696992601	QETVCEASL	NAPD			
D pulax E9HKV0	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
D purpureum P0ZAN5	KETSIHLEK	GFDFLGFN			
M mulatta H9H4M8	AVEDLMHLFF	LCDFACNCH			
M truncatula XP_003629388	PCEDSNHLFF	MCESMLYH			
O sativa AAC13524	DDTVEHVFL	FCDFAACIK			
F vesca XP_004310201	QETVCEASL	NAPD			
I ricinus V5HHM5	QETVCEASL	NAPD			
AT4G04650	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
A thaliana AAD12028	KETSIHLEK	GFDFLGFN			
M truncatula XP_003611652	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
C appendiculata U5HF69	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	
A thaliana AAP79665	KETVVHLSL	GVDFLGFNVG	HFHOKLIID	SRDALQ	

Suppl. Fig. 8.

Alignment: C:\Users\yizuna\Desktop\thesis\filogenias articulo\origen zf-RVT\DominioRT_R_zfRVT_R
Seaview [blocks=10 fontsize=10 A4] on Mon May 23 21:32:33 2016

[illegible]