

# 1 Supplemental Material

2 **Title: Widespread Horizontal Gene Transfer from Double-stranded RNA**

3 **Viruses to Eukaryotic Nuclear Genomes**

4 Huiquan Liu<sup>1,2</sup>, Yanping Fu<sup>2</sup>, Daohong Jiang<sup>1,2 \*</sup>, Guoqing Li<sup>1,2</sup>, Jiatao Xie<sup>2</sup>, Jiasen Cheng<sup>2</sup>,

5 Youliang Peng<sup>3</sup>, Said A. Ghabrial<sup>4</sup>, and Xianhong Yi<sup>2</sup>

6

7 1, State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan  
8 430070, Hubei Province, P R China

9 2, The Provincial Key Lab of Plant Pathology of Hubei Province, College of Plant Science and  
10 Technology, Huazhong Agricultural University, Wuhan, 430070, Hubei Province, P R China

11 3, Department of Plant Pathology, China Agricultural University, Yuanming-Yuan West Road No.  
12 2, Haidian District, 100193 Beijing, People's Republic of China

13 4, Department of Plant Pathology, University of Kentucky, 201F Plant Science Building, 1405  
14 Veterans Drive, University of Kentucky, Lexington, KY 40546-0312, USA

15

16

17

18 **Table S1. Summary of candidates for non-retroviral integrated RNA viruses (NIRVs)**

19

Database	Organism	Accession no. and position (length, bp)	Best matched virus <sup>a</sup>	Related gene	aa identity	E value <sup>a</sup>	Adjacent TE <sup>b</sup>	ORF disruption <sup>c</sup>
Eukaryota genomic	Gastropods <i>Aplysia californica</i> (California sea hare)	GL006980.1: 51763..52047(285)	Cryptosporidium parvum virus 1	CP	64%	3e-24	DNA/Sola, NonLTR/Nimb	internal stop codons
GSS	Plants <i>Brassica oleracea</i> (Wild mustard) <i>Solanum tuberosum</i> ( <i>Potato</i> )	BH939664: 32..693(662) EI814115: 1..138(138)	Carrot cryptic virus Raphanus sativus cryptic virus 1	CP RasR6-3(CP)	45% 52%	6e-47 5e-05		frameshift
GSS	<i>Vigna unguiculata</i> (Cowpea)	EI930635: 3..218(216)	Fragaria chiloensis cryptic virus	RNA3(CP)	41%	1e-05		
HTGS	<i>Lotus japonicus</i> (Lotus)	AP010106.1: 56301..56798(498)	Rose cryptic virus 1	RNA3 (CP)	47%	4e-17	NonLTR/L1	internal stop codon
Eukaryota genomic	Fungi <i>Vanderwaltozyma polyspora</i> DSM 70294(Budding yeast)	AAZN01000409.1: 77491..77733(243) †	Penicillium stoloniferum virus F	CP	35%	2e-05		
Eukaryotic genomic	<i>Penicillium marneffei</i> ATCC 18224 (Penicillium)	ABAR01000142.1: 119..2156(2038)*	Saccharomyces cerevisiae virus L-A	CP	39%	1e-109		
Eukaryotic genomic		ABAR01000272.1:1..1099(1099)	Saccharomyces cerevisiae virus L-A	RdRp	52%	1e-71		
Eukaryotic genomic	<i>Candida parapsilosis</i> CDC317 (Budding yeast)	CABE01000013.1: 736306..738291(1986)*	Saccharomyces cerevisiae virus L-A	CP	35%	1e-99		
Refseq_genomic	<i>Schizosaccharomyces pombe</i> (Fission yeast)	NC_003421.2 37110..38374(1265)	chrIII: Saccharomyces cerevisiae virus L-BC	CP	23%	3e-18	DNA/Polinto	internal stop codons, frameshift
GSS	Nematode <i>Strongyloides ratti</i>	CZ541007 CZ541027	Magnaporthe oryzae virus 2 Leishmania virus 1 - 1	RdRp RNA RdRp	35% 32%	7e-05 5e-16		internal stop codons internal stop codons

GSS	Protozoa	<i>Entamoeba terrapinae</i> M	AM686088; AM680125	Saccharomyces cerevisiae virus L-A	RdRp	61%	3e-61
GSS			AM668999	Saccharomyces cerevisiae virus L-A	RdRp	68%	4e-27
GSS			AM686087; AM667464	Saccharomyces cerevisiae virus L-A	CP	42%	4e-47
GSS			AM678833	Saccharomyces cerevisiae virus L-A	CP	50%	2e-53
GSS			AM670785	Saccharomyces cerevisiae virus L-A	CP	47%	2e-32
GSS			AM675473	Saccharomyces cerevisiae virus L-A	RdRp	38%	2e-36
GSS			AM677307; AM676129	Saccharomyces cerevisiae virus L-A	CP	33%	4e-25

20 <sup>a</sup> Best matched virus and E value were generated by using totivirus-related sequences as BLASTx queries against the non-redundant (NR) protein  
 21 database.

22 <sup>b</sup> The class/subclass of transposable element (TE) or repeat is indicated: DNA, DNA transposon; LTR, LTR retrotransposon; Non-LTR  
 23 retrotransposon. See genome map of TEs in Figure S5 in the supplemental material.

24 <sup>c</sup> The types of ORF disruption (frameshift, stop codon, etc) are listed.

25 † This totivirus-related sequence has been reported by Frank and Wolfe (17).

26 \* The indicated two totivirus-related sequences have been reported by Taylor and Brueen (60).

27 CP, capsid protein; RdRp, RNA-dependent RNA polymerase; aa, amino acid.

28 **Table S2. List of genomic survey sequences which were assembled to generate the consensus**  
 29 **sequences (contigs).**

Organism	Consensus sequence	No.	Accession no. of genomic survey sequences (GSSs)					
<i>Nicotiana tabacum</i>	Contig-1	17	ET873515	ET934781	ET934863	ET967037		
			ET967098	FH297731	FH297797	FH339569		
			FH339650	FH390292	FH390372	FH504911		
			FH504985	FH635597	FH639246	FH642791		
<i>Zea mays</i>	Contig-2	7	FH642879					
			ET940016	FH002080	FH157022	FH894314	FH001811	
			FH157103					
<i>Medicago truncatula</i>	Contig-3	12	FH001736					
			FH208174	FI048635	FI002313	FH232372	FH208234	
			FH687998					
<i>Strongyloides ratti</i>	Contig-4	9	FI009877	ET709315	FH396656	FH339993	FH543953	
			FH687920					
			FH984307	FH939624	ET766028	FH124202	FH610390	
	Contig-1	5	FH215682					
			FH939683	FH463206	FH610469			
			CL993907	CC669533	CG156006	CG437968	CG080070	
	Contig-1	4	CR296971	CR499030	CG966070	CG969600		
			CZ547029	CZ547019	CZ540847	CZ539206		

30  
 31 Genomic survey sequences (GSSs) sharing an alignment of at least 97% identity and 50 bp overlap  
 32 are assembled to generate the consensus sequences (contigs) by running the CAP3 Sequence  
 33 Assembly Program (Huang and Madan. Genome Res.1999, 9:868-877).  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41

42 **Figure S1. Multiple alignment of *ILR2* of *Arabidopsis thaliana*, *gem* of *Festuca pratensis* and**

43 **capsid protein gene of SsPV-S.** The DNA alignment was derived from aligned amino acid

44 sequences. The region of predicted *ILR2* intron is marked with a red rectangular box. *F.p\_gem*, the

45 *gem* of *Festuca pratensis*; *SsPV-S\_CP*, capsid protein gene of *Sclerotinia sclerotiorum* partitivirus

46 S.

47 **Figure S2. Schematic representation of regions with similar sequences for the junctions**

48 **between some non-retroviral integrated RNA viruses (NIRVs) and host sequences.** Colored

49 rectangular boxes with arrowheads indicate open reading frames (ORFs). Green rectangular boxes

50 indicate repeated sequences in eukaryotic genomes detected by BLASTn searches. Blue arrows

51 indicate primers which were used to amplify the junctions between NIRVs and host sequences. Red

52 arrows indicate primers which were used to detect the transcripts of viral related homologs by

53 RT-PCR. Gray sectors connect corresponding homologous regions and the % nt identity is indicated.

54 Red bars represent the matched regions of sequences that are identified by BLASTn; the sequence

55 types and % nt identity are indicated: GSS, genomic survey sequences; EST, expressed sequence

56 tags, WGSs, whole-genome shotgun sequences.

57 **Figure S3. Multiple alignments of the amino acid sequences of some non-retroviral integrated**

58 **RNA viruses and their related viruses.** The regions of RdRp conserved motifs of partitiviruses (A)

59 or totiviruses (B) are marked with red solid-line boxes. FsV-1, *Fusarium solani virus 1*

60 (NP\_624350); RNPV1-W8, *Rosellinia necatrix partitivirus 1-W8* (YP\_392480); VfPV-1, *Vicia*

61 *faba partitivirus 1* (ABJ99996); Dh, *Debaryomyces hansenii* CBS767; Ps, *Pichia stipitis* CBS 6054;

62 Et, *Entamoeba terrapinae* M; Cp, *Candida parapsilosis* CDC317; Pm, *Penicillium marneffei* ATCC

63 18224; ScV-L-A, *Saccharomyces cerevisiae virus L-A* (NP\_620495).

64 **Figure S4. Binary rooted trees of some non-retroviral integrated RNA viruses (NIRVs) (A, C)**

65 **and their related viruses (B, D) with Ka/Ks ratios plotted on branches.** These trees were

66 obtained by a neighbor-joining analysis of a codon sequence alignment under the Maximum

67 Composite Likelihood substitution model. Branch lengths are not to scale. (A) and (B), rooted tree

68 of capsid protein-like sequences of NIRVs and partitiviruses respectively. (C) and (D), rooted tree  
69 of RNA-dependent RNA polymerase-like sequences NIRVs and partitiviruses respectively. A ratio  
70 of less than 1 indicates purifying selection to conserve protein sequence. Note that the Ka/Ks ratios  
71 are averaged over sites.

72 **Figure S5. Schematic representation of endogenous non-retroviral dsRNA-like elements and**  
73 **flanking transposable elements in eukaryotic genomes.** Yellow rectangular boxes indicate  
74 non-retroviral integrated RNA viruses: arrowheads within boxes indicate sequences retaining the  
75 reading frames of viral genes; ψ within boxes indicate sequences containing frameshifts and stop  
76 codons compared with viral genes. Green rectangular boxes indicate repetitive or transposable  
77 elements annotated by BLAST. Blue rectangular boxes indicate repeats annotated and classified by  
78 Censor. The Censor server automatically classifies all known repeats and adds the classification to  
79 the report. The class/subclass of repeat is indicated: DNA, DNA transposon; LTR, LTR  
80 retrotransposon; Non-LTR, non-LTR retrotransposon. CP, capsid protein; RdRp, RNA-dependent  
81 RNA polymerase. \* Two genomic contigs of *Ixodes scapularis*, which are homologous to each  
82 other, one contig (ABJB010791923) contains the viral homologous sequence whereas the other  
83 (ABJB010911717) does not. Gray sectors connect corresponding homologous regions.

84  
85

**Figure S1**

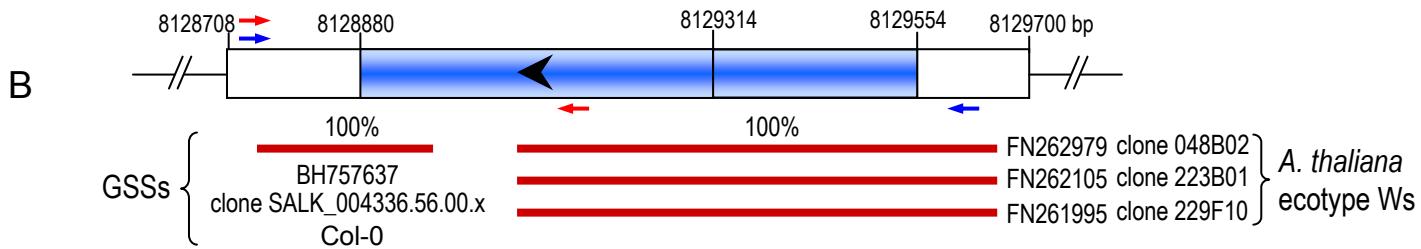
<i>ILR2</i>	1 ATGGCCTCTGAATCTCAACTCACAGACGAAGCTT	GAGAAG 42
<i>F.p_gem</i>	1 ATGCGTCAAGAACCAATGCCTCATTGAGAGGCCCTCGCTCGAGCCAAAGCTGCCGTCGGAA	CTGAAG 93
<i>SsPV-S_CP</i>	1 ATGTCTTCCCGTCCGCTCGAAGGTCTCACTCAGGAGAAAACCTCT	CTGGAAAGAACCGTGCCTCCAAGAGTCTGGAAAGAAATCTGCCCCCCCAGTATTCTGTGATACTCCATCCGAA 132
<i>ILR2</i>	43 AAGACTTGTCTTAGTTCTT	CTT 63
<i>F.p_gem</i>	94 CAAACAGGACTGACCTCGCTCGACGAGCTTGACCCCCAAGCAGTTCGACGATTGGCC	CCCTATGCTCCAACGAAGCTGAAGCCGTCACCTCGGAGCCCCAAACCTCGAAGAGGCTCACACGC 219
<i>SsPV-S_CP</i>	133 CAATCGGATGTTGAATCGGACATCCTCTCGAACTCGAAGTCTGACGCCATGACGCCCTCCCGTCCCCTCAAAGTCGAAGAAAGGTACAACACGCAACAGAAGAAGGAAAATCTGTCGCCTCC 267	
<i>ILR2</i>	64 TCTAGTTACCCGAAGCAAATCGTGTACCCAAACGCAAGGTATCTCACACTACTACCCCTCCTGCCAC	132
<i>F.p_gem</i>	220 TCTGACGTCGACACCTCCGATGACTACACCCCTGACTCCGCTCTGACATG	CTCATGCCTTATTAGGCCAAACCTTAGGTATGTCCTCC 345
<i>SsPV-S_CP</i>	268 TCCGCTGGATCATCAATCCACCTACATGATGATGCTACCGCCCTCAACTTTTCTCACCGAAATCGAA	CATGTCACCATCTCAACCTATACTCAAGCTGTTGG 378
<i>ILR2</i>	133 ATGATG-GTCCATGCTATGAAACATTACACTTCCGATAATTTGATTCAAACGGGCAATCCCACATCACCCCTACATCCTCCGCTCTACTGTGGTTCTCTTTGGATCCAGTGTCTA	GA 258
<i>F.p_gem</i>	346 ATGATGGACTACATCTCCACCTCATCAACGACAATTATCGACAATTCTATTTCAAGAACGCTGTCGGACTACACCCCTACATCCTCCGCTCTACTATGGCGTATTATTGGATCCAGTGCCTCCCG 480	
<i>SsPV-S_CP</i>	379 TC GAT GT AC GCG T CTC GAC GCT AT GC AT GAT CT CG TT GT GATA AC GCT T CTC GTC GCT T CT GTC CC T ACT AT CA CG T CG CC CT GT CC A ATA T T A C T AT GGT ATT CTT CAT ATT CA AGT CCT CG GT 513	
<i>ILR2</i>	259 GCTGGAAACGATGTGAATGATCTTA C T G T C A A C A C C G G T T C C T A A C C G G T T C T T G G A C A A C C A T C C C T T G G A A A C T C T C G C T G T T C C C G G C C T C C T C G G A C T C T T C A A G A C T C T T G C T C T T C G C A G 393	
<i>F.p_gem</i>	481 GCTGCCCGT GACGTCCCTGTCCTTGAGGACCAAGATATCAGTTTGTAGATGCTACCCCTCAAGTCGCTTACCATCTCCAGCCCTTGTGGTATATTCAAAACTTTGTGCTCCCTCCAG 615	
<i>SsPV-S_CP</i>	514 G C C A A T C A A G T C C C A A C A A T C T C T C A A G C T G A T T T C A G T T T C T G C G T T T T G A T C G A A T T T G C T C T C G A A G A C T C C T G T C G C C G T C C C T C G T C T C T T T C A G A A C C T C G C C G T T C A A A 648	
<i>ILR2</i>	394 CCGGAGTTTCTCATATCATATGGAAA GT GTATCCTCGATCCCTGCCAACCA G C C C C G G T C G A C G T G A C G C G T T C A T G C G A G A T C T T C T G C C T A A C G T C C C T G G C A T C T C G C C 528	
<i>F.p_gem</i>	616 CCAGAGTTCCCCACC-TACGGTAAGTTTACCGCGCTCCCTCACTCCCCGGCCACCCGCGCGACGAGTTCATGCATGAACACGTGCAAAATTCTTTGTCCCAACATCCCGGGATCATGCC 744	
<i>SsPV-S_CP</i>	649 CCTGATGGAAATCGT-TTTAACCTGGTTGTCCTCACTACAACAAAC-TACGGTCCCGAACAGGTGCTAAATCACCCCGTGTGAACGCAACATCACCGCACCTCCCCACTACCTCAGATGATCTCA 774	
<i>ILR2</i>	529 CTCTTGGAAAGACTTACCGCCTTTTACCAATACCTCCGCTCTATCCAAAGAGAGGGAG-GCACATTCCCGTTACCCAAATACCCTC	TCTAAGTCGGTACAAGACCTTTGGCT 645
<i>F.p_gem</i>	745 CTCCCTGCCAACCTAACAGTATCATCAACCCGCGACCTCGCAACGCTGTGTTCCCGAAAGGGAAACACATCCCT-GTGACCGCCACGCCGCCCCAGGCCACTGTTTTGCCACCATCTTCCCGAT 878	
<i>SsPV-S_CP</i>	775 CT CCT TAATCTCTTGGTCT - T C C A A C G T G C C T C C T A C A G C T A T G G A T - - A C C G C A G G A C A A T G G G A C C A - T T C A C T T T C - - G C G C T G G T G G A A C C A T G C T G G T T G C T T A T G G T G C T G G 896	
<i>ILR2</i>	646 CCATGCAAGCAGGACAGGGCTGAAAATGGTCACTCGTTACACTGGCTTACAGTACCCCTGTGAAGCCGATCAAAGCCTGAATGAGGCGTTTGTGATTGCTACAGCAATTAACTTTAGGTCACTCTGC 780	
<i>F.p_gem</i>	879 CGCTGCAAGCAGCAGCAATTGGAAAGTGGAGCCTGTGCTCTCTCTGGCTTCCAGTACCGGTGCGAAGCCGATCAAAGCCTTAATGAAACTTTGCTGAGCGATATGAATGTTCAACTTCCGCCACATTGGC 1013	
<i>SsPV-S_CP</i>	897 ATTGCTACTGATGCTGTCATCGGACCAACCTTCGGCCCCAGGCAATTACGTTCTGGACACATGACACCGCAATTCAATAGGAAACCTGTCCTATCAGTCGAAGAATGAGGCTTCAACAAATGAACCG 1031	
<i>ILR2</i>	781 CGCTGATAACCTCGAAAGATCTTCCCTTCCACATGAAACACAGCACTGGCTTGGTTCAATCAAGTAAAAGAGTTGGATGATGTTGCTGCATCCTTGAAGGATCCGGACCCCTGCTGACTGCTCTC 915	
<i>F.p_gem</i>	1014 AACGGATGATTTACGATTTCTCAAGCTTTGAGTGAACGGTGACCTTCCCTGGTCCCGACATGTCCTCCGATGCCGAAGCGCAGCCGTTAACGGCTCTGGCACATTAGTGACTGTCCTCC 1148	
<i>SsPV-S_CP</i>	1032 TGCTAAAGATTCC- - - CCACGCACTTACTGTGGCTAGACGGTCAATTTCACGGTTTACAAAGCCATTGAGCTGTCAGTCAGAAAGCTAAGTATTGCGTGGATCTACTACACTCGCTAATATCAATTCC 1160	
<i>ILR2</i>	916 ACATGGGTTGGTGTAAACCAAGTCATGGCTCTCTACCCCGGAAACATCTCCAGAAATCACCATTGTATTGCTGATAAGCGTGCAACCTATGAGTTGGTACCAAGCTCAAAGCACAGTTG 1044	
<i>F.p_gem</i>	1149 AACCGGGATAGCTCAACCAAGATTGCGTTGAGTACATGCCCGCCCGTACCGACGTAACCGCCCCCAACACAAACGCCGACCCCTCTCGCTTCCCATTCACCATCAAGCTCCAT 1277	
<i>SsPV-S_CP</i>	1161 TTCAACCGGTCCTCCGCGTGTGAAACTACTGTGCGTGTGCCCCGGCGCTGCTATTCTCAAGCAACCGCC-TGGTATGAAAGTTATCCCTCAATCTAACCTAACGTTAACACTCGCGTCCG 1286	
<i>ILR2</i>	1045 CAACCTCCCTCCCTGGCAAGGACTTGGCCCTTCTCAGACATATATCAGGATGTTCTAACCATCCCTCTTGGCGACGTTGGCTGCAAACCTCTTGTATCAT 1173	
<i>F.p_gem</i>	1278 GAACCTTCTGCGTATCACAAAGTGTGGCTGACCGCCAGACCCACATCCGATGTTCCGACCCACCGTATTGGCGAAGTGGAAATCATCGTGGCTGC 1403	
<i>SsPV-S_CP</i>	1287 TTCCATCTGAAACAGATGTAAGATTGGTACCTCGCTGCTACTTGT-AATATTACAGGTGGAAACTTCACCCCTGGACCGTTCGTCGCTGGGATTTTCGCT 1415	
<i>ILR2</i>	1174 GCCTATTGGCTCCAGCCTTACCGAT-AACAGCTCTAACCTCACTATCCCACGATGTCAGCAAGCAAGCATTCAAAGGCCAACGCCACG 1261	
<i>F.p_gem</i>	1404 CCCCATCGAGTCGTCCTCCAGAC- - - GAGTCTTCTAACCTCTCTACAAAGGAACTCGTGCAGGAGCTGCTCAAGCCCAA 1485	
<i>SsPV-S_CP</i>	1416 TCCTATTAACTCGAAAAATCCAACCGAAGTTACGCAATCGCTATGGCTCAACCACCATGACGCTTAGCCTCAACGCCCTAAAGGTGGAGAAGCTACTGATGAT 1527	

Figure S2

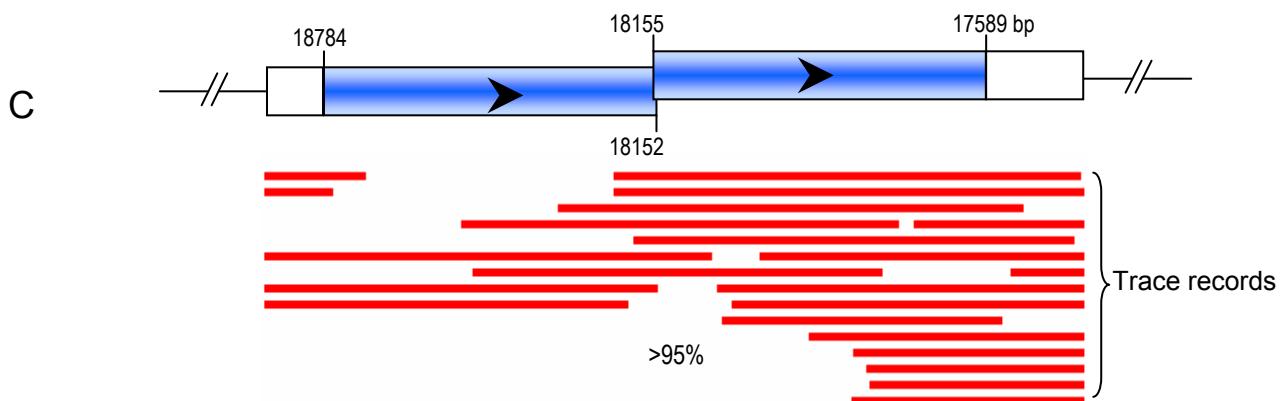
*Arabidopsis thaliana* ecotype Col-0 chr 3 NC\_003074.8



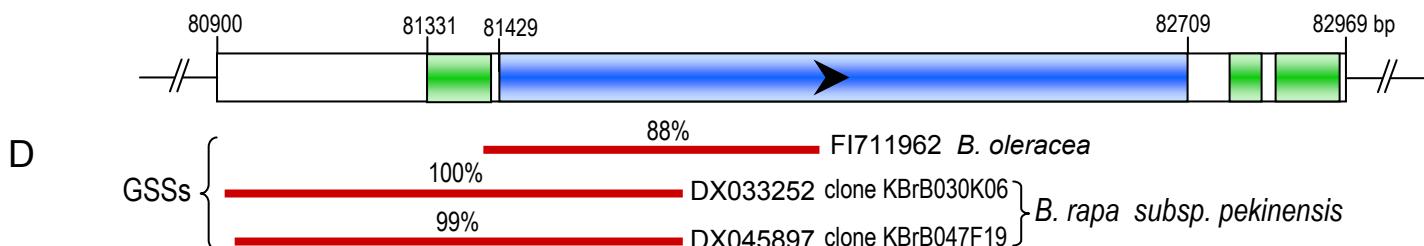
*Arabidopsis thaliana* ecotype Col-0 chr 4 NC\_003075.7



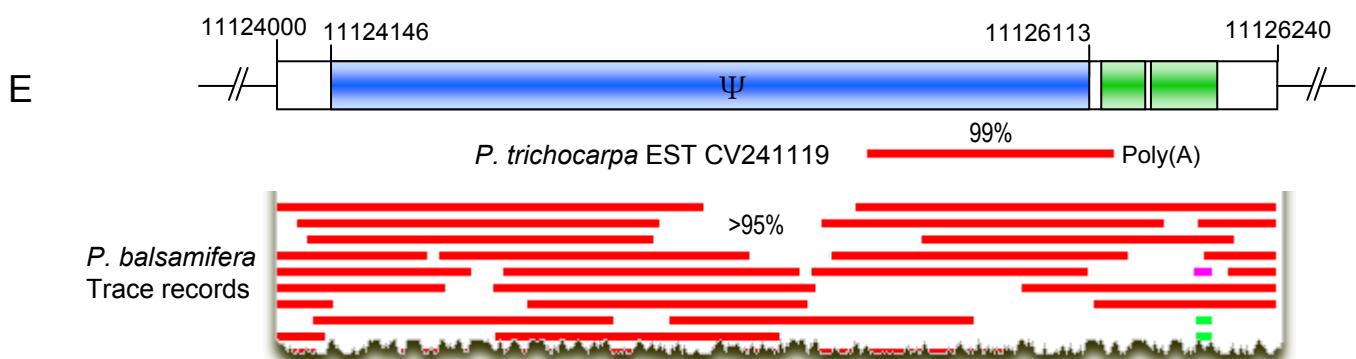
*Acyrthosiphon pisum* strain LSR1 NW\_001917032.1



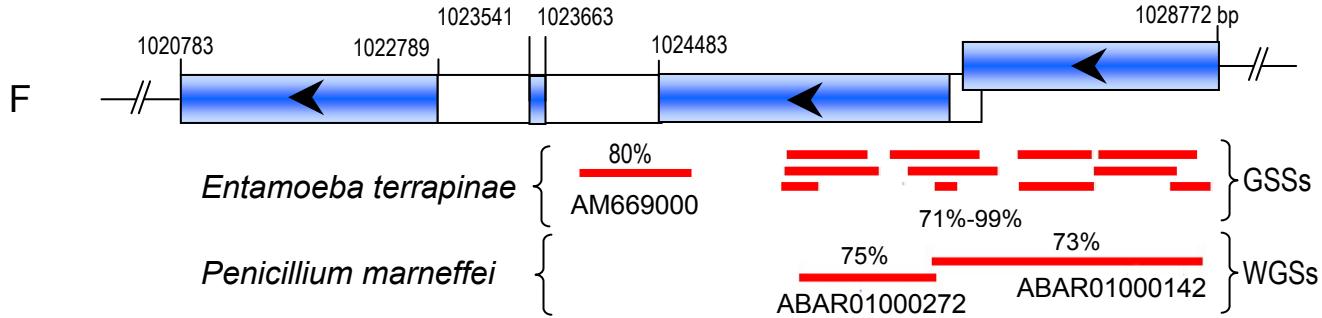
*Brassica rapa* subsp. *pekinensis* clone KBrB070J05 AC189442.2



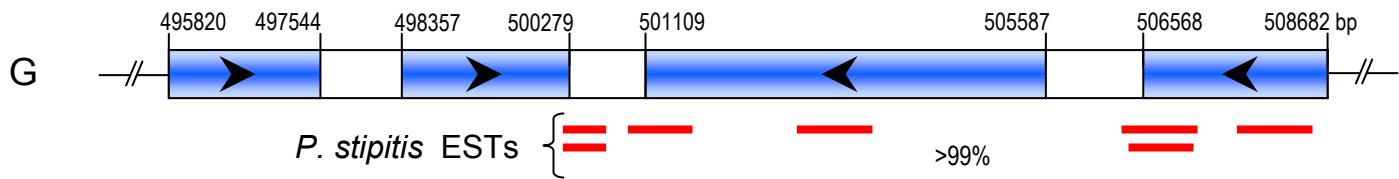
*Populus trichocarpa* NC\_008473.1



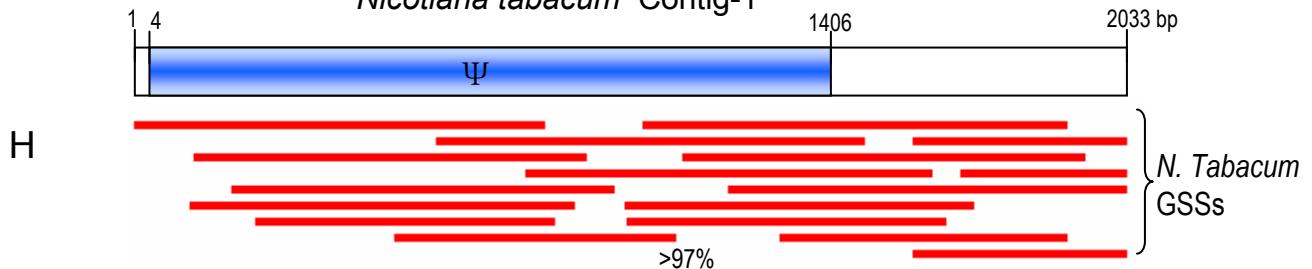
*Debaryomyces hansenii* CBS767 chr B NC\_006044.1



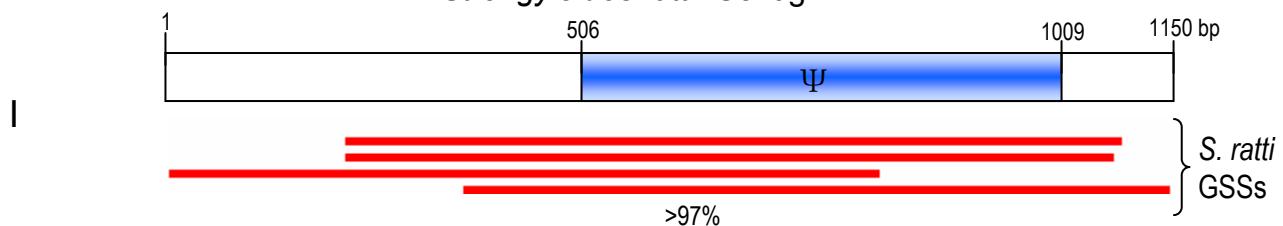
*Pichia stipitis* CBS 6054 chr 7 NC\_009047.1



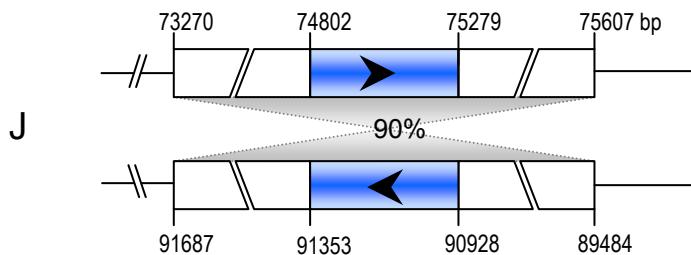
*Nicotiana tabacum* Contig-1



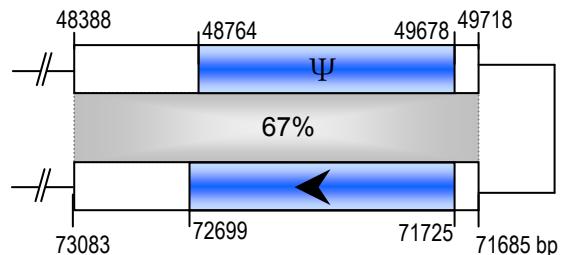
*Strongyloides ratti* Contig-1



*Medicago truncatula* chr7 AC175047.3



*Lotus japonicus* chr1 AP007812.2



# Figure S3

FsV-1\_RdRp  
*D.grimshawi\_AAPT01020539*  
*R.prolixus\_ACPB01059984*  
*A.pisum\_NW\_001917032*  
*RNPV1-W8\_RdRp*  
*VfPV-1\_RdRp*  
*A.pisum\_NW\_001916518*

1 -----MVELFVMVDPTTKRRRIQSTLGPFLSVPGLQEIARYGGYATYRAVRNTPWI RQLSKLFDPD-LYGNIYGFTRRP----- AGPE 78  
1 -----MDIGKAEGFEPPETSPYLTKL YPSCKEFTQSVALKLIRDDIKEYRVNEAKRP----- 54  
1 -----MEYH-RQYAPATVRNDP----- 16  
1 -----MPSSP----- 5  
1 MVLTIIRDYLHEAQLRLKKEWQTFQKSDQESGYSKDLPTDYDLRYYDSARDYDAEKHKTEYQHNFA LTHE-RYTQMNADRNEPEFYRPLEDNELPDIRFPAPGI TVLPFRYHTGQIVETTDEL PDSG 129  
1 -----MKNITVI GNL PNL AKP-I YGGSTADTNPAYAETVDHALYKFLSPEEAELVI NGYRRSP----- VNDEALKDDIAKLDSPY 74  
1 ----- 1

FsV-1\_RdRp  
*D.grimshawi\_AAPT01020539*  
*R.prolixus\_ACPB01059984*  
*A.pisum\_NW\_001917032*  
*RNPV1-W8\_RdRp*  
*VfPV-1\_RdRp*  
*A.pisum\_NW\_001916518*

79 GMYKSLMKFGESMPRTDMSTVQRSAAMTAITAARKRFKTPVKFEPLESEVGQ----- HMRRTDSAGVSFPGKKKGDVMERIY-AEG 160  
55 -VDPLIDEAIRRTAIAFRLPQREAIIHLNDVFGQDLPIW-SSSPGI PWSQI GY----- KTKDDIRKDPE-AVQ 119  
17 -VYAMVLQSVRQ-----EFIPRDKIIPLTLGAVEKYPNPKDRSPGLPWKGKGY----- KTKGDVLADPV-ACA 78  
6 -----TPGFFFFLSKNQFLKKAQFADKFSTDF----- 32  
130 FSLHPLIDYLTKTKWLYHRYPI DKYCRPLGTTNATFSDFNREQI PSAPI DETHRKNMVLPLVIYFLNALPFLPIHFVDTRFCGTPKHTATGYFQRFSTFRTHAYYARNKL YALRPTSKGYFFNTVYEFSR 259  
75 FQVTKDEHYYKAIAHVKKLFTPDTPLKPVHFA DLRHYPWQLSSNI GAPFATSPE----- WQDYINNKFLTGQTAENYRDLFKEAHHTDL YPKMLDRRMTKRNLYNEMF-LIN 180  
1 ----- 3 4  
161 RWLGHRMKQG-----GKGRFDPRKVRMPPCLATQRGHLSPRDP-KTRLAWIYPSEMLMV EGYAPTMYKAFE-----AMPDSPLLG-----KGSHRLFSEWSAATPG-MRL YG-----DFSSFDTKVPA 271  
120 RVRVFWHQVK-----ARKDISLPDCCAYMRFHLCKVGD-----KVRAVWGYPATVTFAEAMFALPLIRAYQ-----KRTTPLAYGYETGTGSMRKIFQQFQG-----SHFLG-----DFSNFDKTPA 224  
79 TWRKRKWRNIG-----RGRSET-----LPDVALFVRAQVTGIDRE-KVRAVWGYPIDV-----VEDGRSRFYPYL SWLK-----KTSAPIGYCVEVATGGMSYI NDMLTAHPD-----AKYLITDWSCDFDKTP 186  
33 -----IHTYPTIMFTRNHVSEERNGVLKNRPVYAVDELFLTL ECMLTFPLHLQAR-----QMSSCI MYSLETMRGNNKYLDSDIVQQY-----QSYFTADWSSFYQTIPF 125  
260 TWMHHIKEHGYPFVPSHDALDNARQYRIFM QKHVMTL FTRNHISDRDGFLLKQRPVYAVDDFFLCELMISFPLHVMARYPINGIKSCIMYSFETI RGSNRYLDSIARDF-----ISFFTIDWS SFDQRVPR 385  
181 RKSIIHIIKNG-----RTTNDHGHDLLKYWNTAFARQHLEDPPDKVRLVFGAPSTLLMAELMFIWPLQVSLL-----ARGPQSPMLWGYETLTGGWSRLFSWASDTMPRYSTVTLTDWSRFDKDARH 297  
1 ----- 5 6  
1 ----- 7 8  
272 WLHTAFDIL-----HDNI DWLHWRGKPTTKRSRQWKNVWDGM/VYYFINTPI LMPDGRMFRKR-RGVPSSGWSWTQLVDSVVN-----WILVKYLSLCQGVNAKLRVLGDDSAFMAETMDL 383  
225 WLKLAFDVM-----AINLDFTKYHD-HGVPCTR-SIIRVFETLEDYCINTKIRMCNGERYLKR-SCLASGSYFTQLVGSISN-----HILLQYAALVKKVEIQDI VVFGDDSLMGTRSLTLP 334  
187 WLIRDAFQIV-----YDSMDWHSVVKDSEG-----KVW-----VCQPQRLLL----- 222  
126 KLVRTYCHDFIPSLLIVDHGYHSTYEYPNHPSTSDDKMAL-VMNTLEFLFYFWYKNMFLSQDGFAYQRLYACI PSMQYIAIYL SMI-----HAMIEFDIPEHEIEEFKFLIMGDDNSAFTQWSLTKT 248  
386 VITDIFWTDFLRQLIVINHGQPTYEYPAYPLSEHDLYK-RMNNLLHFLHTWYNNMVFVTADGFAYL RSAAGVPSGLLNTQYLDSCFCNLFLIDGLIEFGFTQAEILSIVFFIMGDDNSGFTMMDI ERL 514  
298 TVIIRDIIAMI MRPMFTFTEGYHPTISYPNAPDTNP-----RL ENLVNWM TDAILTVPLMLPDGQLLQFNHSGI YSGYFQTQLLDSIYNCVMIYTILSRMGF-----DLDKFCIKVQGDDSLMISPONYTL 419  
44 SVDNINHNVIKSYIDFENGFPVTFDYPETSTEPQ-----RLLNWLNLTYDAIKHTPDVI PNEDTYVRLVGLASGFQTQLLDSLYNMSILLTVLARMGI-----NIESVKMKVQGDDSVIALTELIPPL 163  
FsV-1\_RdRp  
*D.grimshawi\_AAPT01020539*  
*R.prolixus\_ACPB01059984*  
*A.pisum\_NW\_001917032*  
*RNPV1-W8\_RdRp*  
*VfPV-1\_RdRp*  
*A.pisum\_NW\_001916518*

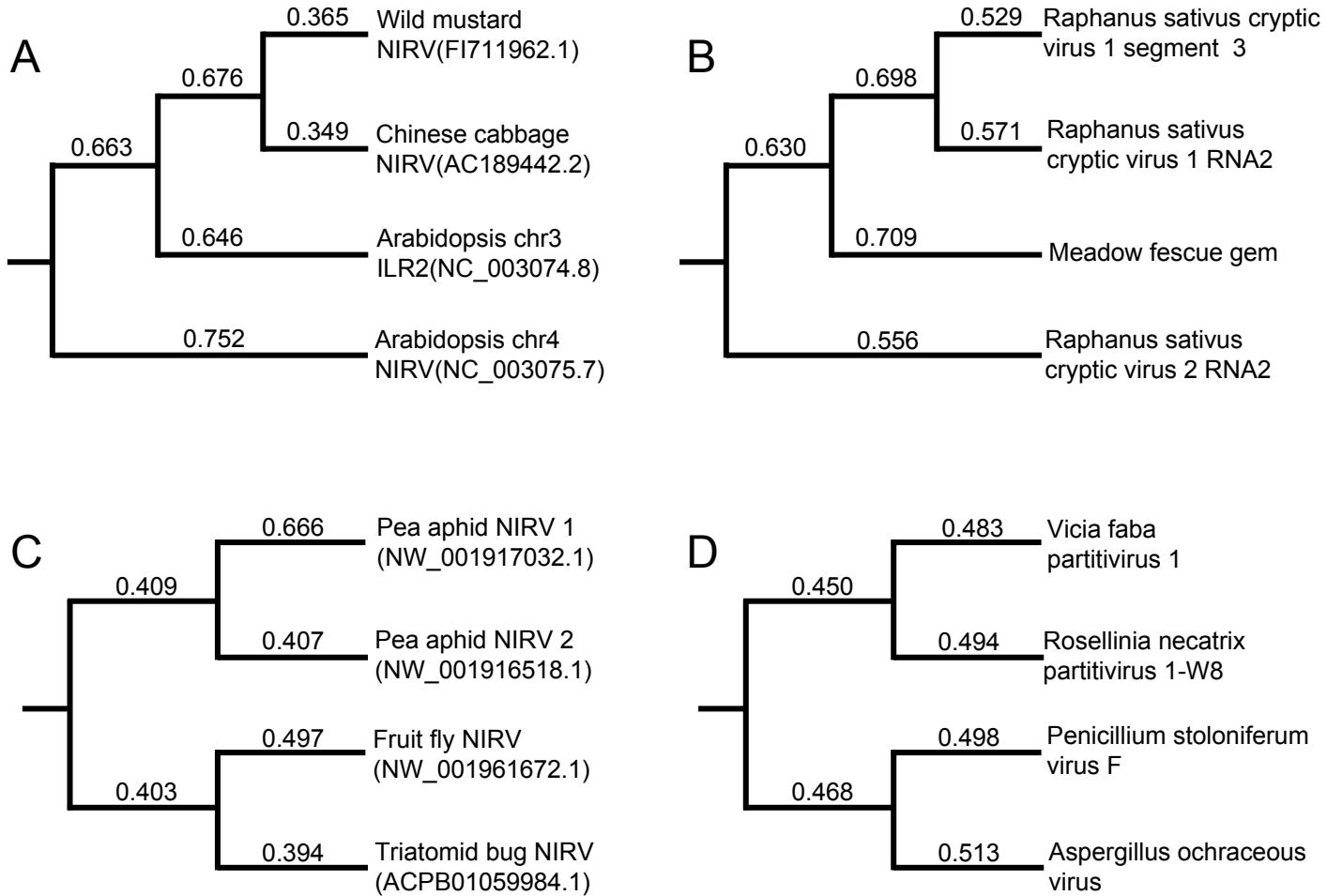
384 -----SVAEAEDAAVGMDLSDEKSI SVEDATELKLLGVRYRDGHAFRETEEWFKLALYPEGDPDIATSLTRLVGLWI GGAMWDTKFSRME-----YFQGCYPCPSEGWF SKDQRRWME I VHGGRAPRG 503  
335 -----DDVQEVLRLSGMNVNVEKSGFSRSVTNLTFGLGYGVNGFPKHHKRFVASLWVPERNDKCWDDVASRALGI I YANL GVDDRIDL----- 418  
223 -----NATRFNFLNFCFLSGTPGNA----- 242  
249 -----QRFVKFLSEFALIKYNMKLNLKTTVTTSERQHQLTSLSYECNCGNPKRPLGKLLAQLCYPEHGYVD-KHMSAGAI GIAYASCAQDYTFYL FCKDVFYTFPPYAEPLTSADVKMRKYL PGVFKL MDEIP 375  
515 -----TQFIEFFESYALKRYNMVLSSKTKSVI TTLRSRIETLSSYQCNGGNPKRPLGKLI AQL CYPEHGPKD-KYMSARAIGIAYAAAAMDEFHEFCRDIIYHTFLPYAAPIDEHTLSMATKHL PGYFKMLDNIA 641  
420 E-PSFLKTFGDIKEYFGSTLNLKKSEI LPSLDGAEVLYKRNHGTMQREEIQLLAML RHPERTASL-PSLMARCI GVTYANCGNFNRVSRICEDIYKFLQAGGFTPDFPGLPGGL RYRQNYI PSY----- 543  
164 LHSSFLKTFSDYAQQYFGAILNITSKINNDLNGLPLLGF TNQACIT-----LLLYCCTES----- 220

FsV-1\_RdRp  
*D.grimshawi\_AAPT01020539*  
*R.prolixus\_ACPB01059984*  
*A.pisum\_NW\_001917032*  
*RNPV1-W8\_RdRp*  
*VfPV-1\_RdRp*  
*A.pisum\_NW\_001916518*

504 WT KKNLNF-----WRSI FYTL----- 519  
-----  
376 TYLTELQFPT-----YSDILMTIILIFF----- 398  
642 SEIKFDSFPTLEMVQDKYSRQGYL SHPKWNDAHFKFLPETYVPPNNIKTMTDYQLEHKL DTPVPHSLF----- 709  
544 VPIDL SHFPT-----YFENVEHLQDPHKPILSEKHWP----- 576  
-----



Figure S4



**Figure S5**

*Arabidopsis thaliana* chr3 NC\_003074.8



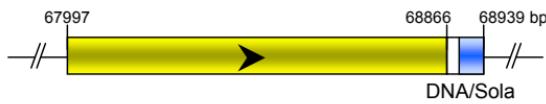
*Arabidopsis thaliana* chr4 NC\_003075.7



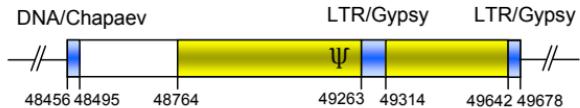
*Brassica rapa* subsp. *pekinensis* AC189442.2



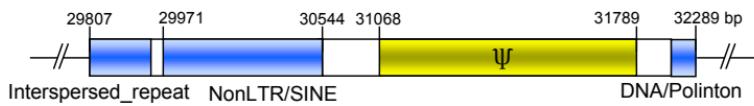
*Medicago truncatula* AC196856.3



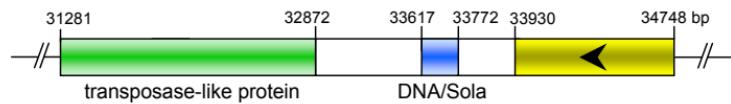
*Lotus japonicus* chr1 AP007812.2



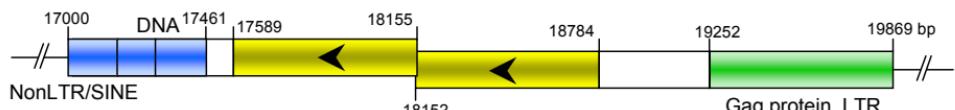
*Entamoeba histolytica* HM-1:IMSS NW\_001915030.1



*Acyrtosiphon pisum* strain LSR1 NW\_001916518.1



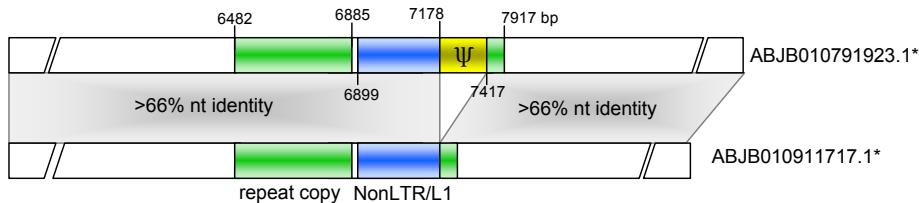
*Acyrtosiphon pisum* strain LSR1 NW\_001917032.1



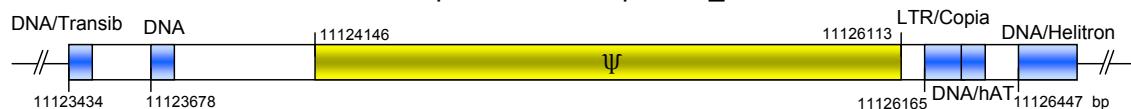
*Drosophila grimshawi* strain TSC #15287-2541.00  
NW\_001961672.1



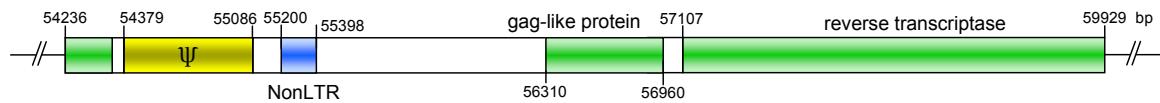
*Ixodes scapularis* strain Wikle



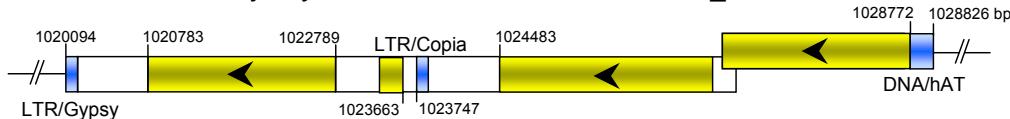
*Populus Trichocarpa* NC\_008473.1



*Aedes aegypti* strain Liverpool AAGE02000678.1



*Debaryomyces hansenii* CBS767 chrB NC\_006044.1



*Pichia stipitis* CBS 6054 chr7 NC\_009047.1

