Supplementary Material

Giant reverse transcriptase-encoding transposable elements at telomeres

Irina R. Arkhipova, Irina A. Yushenova, Fernando Rodriguez

Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA 02543, USA

Supplementary Figure Legends

Fig. S1. Examples of terminal DNA loss and recent *Terminon* transposition in *A. vaga*.
(**A**) Two allelic scaffolds, one of which underwent loss of *CzcO* cyclohexanone monooxygenase (now present only in single copy in the *A. vaga* genome), and was healed by telomeric repeat addition and by *Ath-W* retrotransposition. Truncation occurred within a MITE TE, leaving the thioester reductase intact. The allelic regions, including the common MITE fragment, are 97% identical. (**B**) Telomere M1 (fosmid 184A11 from (Gladyshev and Arkhipova 2007)) compared to scaffold 868 from the same clonal culture four years later (Flot, et al. 2013). Blue arrows, PCR primers initially used to amplify sphingomyelin phosphodiesterase (GDPD), which failed to amplify it in 2010. GNAT, N-acetyltransferase GNAT family. Each separate transposed unit in the head-to-tail array is marked with a bracket. Other notations are as in Fig. 1. (**C**) Regions of HHR-Q homology surrounding the ORFs coding for a TPR protein and a seven-transmembrane receptor (7tmr). An HHR-W2-containing fragment is present in direct orientation.

Fig. S2. Intragenomic coverage of *Terminon* families in the *A. vaga* reference assembly.
Reference scaffolds from Table 1, each diagrammed on the top, were used in BLASTN searches of the *A. vaga* genome assembly scaffolds at NCBI. The graphical overview of the search results is displayed in red. The coding potential for each family is shown in detail in Fig. S6.

Fig. S3. Phylogenetic analysis of *Terminon*-associated enzymatic ORFs.
Unrooted maximum-likelihood (ML) phylogenetic trees were generated using codon-based nucleotide sequence alignments. Clade support values >50% are shown at the nodes; scale bars, nucleotide substitutions per site. Taxon names are as follows: *Av, Adineta vaga*, black; *As, Adineta sp. 11* natural isolate, green; *Pr, Philodina roseola*, red. Asterisks indicate a defect in ORF; 0i-9i, number of introns. Intron gains and losses, when definable, are shown by black and gray triangles, respectively. (**A**) ML tree of catalytically active (ILOM, W, NT) and catalytically inactive (JVXY nc) RT ORFs. Frameshift-containing clades are marked by #. The alignment spans the extended core RT including the C-terminal thumb domain and the conserved N1-N3 motifs at the N-termini (Arkhipova 2006); ORF1 is not included in its entirety, and is presented separately in Fig. S7. (**B**) ML phylogram of GIY-YIG-like ORFs. Clades marked with purple and magenta brackets are both present in their respective families and have been co-evolving with the corresponding RTs. (**C**) ML trees of the full-length Rep-like ORFs (Rep-FL) were split into the N-terminal (HUH-Y2) and C-terminal (S3H) domains, and analyzed together with the respective Rep-N and Rep-C ORFs containing only one of the domains (underlined). The non-catalytic Rep-Nc clade is marked in cyan; /, 5'-truncation. Rep-FL ORFs were assigned to arbitrary clades a-d; number of introns varied from 0 to 7, and is not shown for simplicity.

Fig. S4. Programmed ribosomal frameshifting in *Athena* RTs.
(**A**) Limited sequence conservation in the region spanning the frameshift site in the ILOM and W clades (using MView visualization tool at EMBL-EBI). Note that in the O clade, a shift of the slippery sequence to the left (underlined) has occurred in *A. vaga* (see also panel D). (**B**) Secondary structure-based alignment showing the stem-loop conservation in the W and VX clades. (**C-E**) KnotInFrame *in silico* prediction of -1 frameshift sites, including slippery sequences and simple pseudoknots, in representative *A. vaga* families.

Fig. S5. Graphical representation of pairwise similarities between RT coding sequences from all families.
Each family was plotted as nucleotide (**A**) and amino acid (**B**) sequence to highlight the contrasting degrees of divergence between ORF1 (N-termini) and RT. Representatives from each family (left column) from each of the four clades (right column, square brackets) on reference scaffolds (center column) were compared to their neighbors on the phylogram (Fig. 2) with Easyfig2.2.2. Identifiable cases of intron gains and losses are shown in (A) by yellow and blue triangles, respectively. The yellow box marks the location of the GDSL domain in the I clade; nc, lack of catalytic residues in RT-derived ORFs. The position of the frameshifting pseudoknot is marked by # on the top, and by vertical lines connected by a dotted line in other sequences in (A). The region corresponding to the core RT motifs 1-7 (Xiong and Eickbush 1990) is marked by a square bracket on the top (panel A, dashed line); motif RT5 is shown in white font.

Fig. S6. Structural organization of *Terminon* families.
Pairwise nucleotide sequence similarities are shown for individual scaffolds containing representatives of the W clade (**A**), NT clade (**B-C**), and ILOM clade (**D**). Scaffolds are grouped by RT phylogenetic relatedness (mini-phylograms on the left, with scale bars in nucleotide substitutions per site). Reference scaffold/contig IDs from *A. vaga* and *P. roseola* are shown in red type; *Adineta sp.* contigs (As), in green. Contiguous units from the same family on each scaffold are shown against the yellow background; from other families, against the green background. Blue arrows against the green background denote ORFs from other TE classes; blue arrows with no background, host genes. Other notations are as in Fig. 1.

Fig. S7. Properties of coiled-coil (CC) motif-containing ORFs.
(**A**) Sequence logos of different (gray bar) and similar (blue bar) regions from 67 NWT-like (top) and 46 JVX-like (bottom) ORFs. Light gray arrows mark family-specific intron positions also shown in (B); pink bar, weak homology to HTH motif. (**B**) Prediction of coiled-coil motifs in NWT-like (top) and JVX-like (bottom) ORFs. The gray and blue bars mark the regions shown in detail in (A). Black arrow, intron position conserved in both CC-ORF types; dark gray arrow, intron conserved in all NWT-like ORFs; light gray arrows, family-specific introns in P and J families. (**C**) Maximum likelihood (ML) analysis of NWT-like CC-ORF amino acid sequences, including: stand-alone CC_{NWT} ORFs located upstream (u) or downstream (d) from the catalytically intact RT; ORF1's from frameshifted W clade RTs; and N-terminal moieties from RTs of the NT clade. ORFs from *P. roseola* cosmids (Pr) are in red; from *A. vaga* scaffolds, in black; and from *Adineta sp. 11* (As), in green. Asterisks mark a defect in ORF. (**D**) ML analysis of JVX-like CC-ORFs, including stand-alone CC_{JVX} from NT and W clades; ORF1's from the frameshifted VX clade; and N-terminal moieties from the non-frameshifted J clade RT-like amino acid sequences. Scale bars, amino acid substitutions per site.

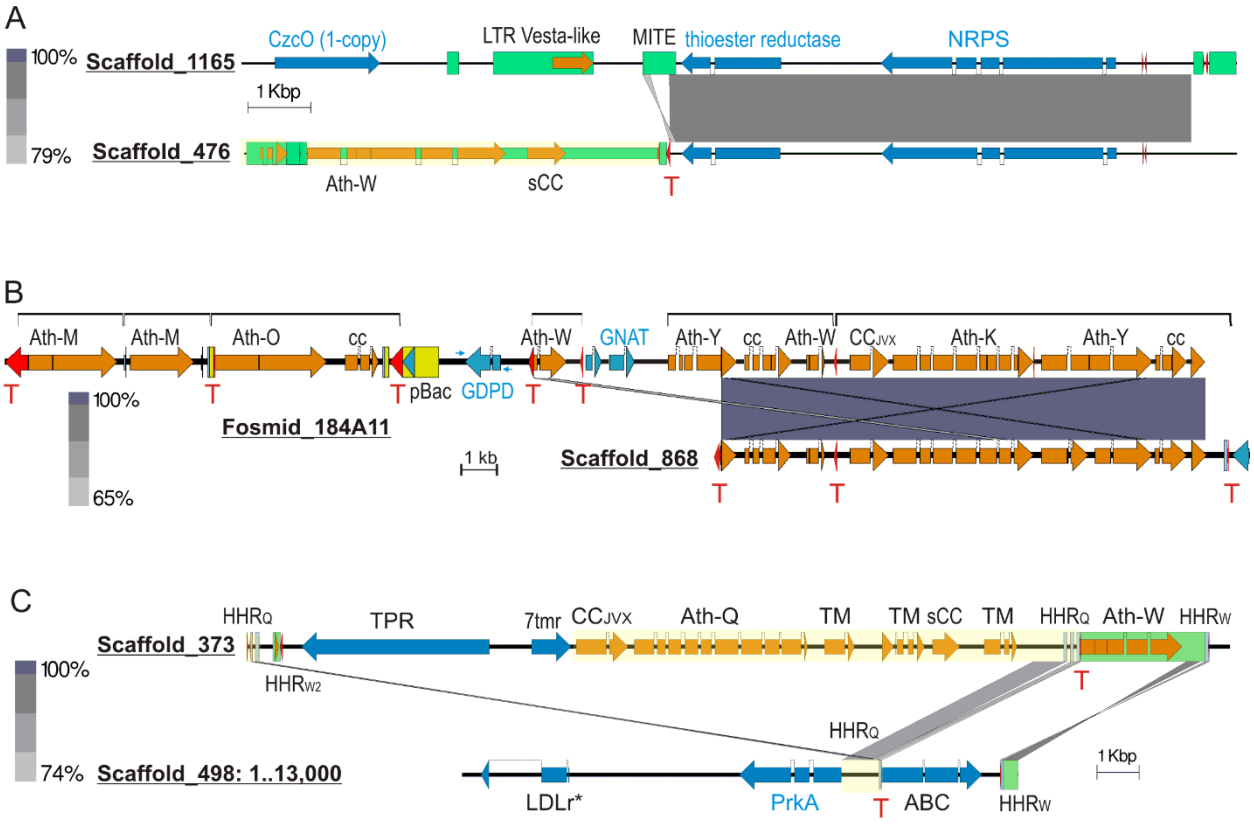

Fig. S1

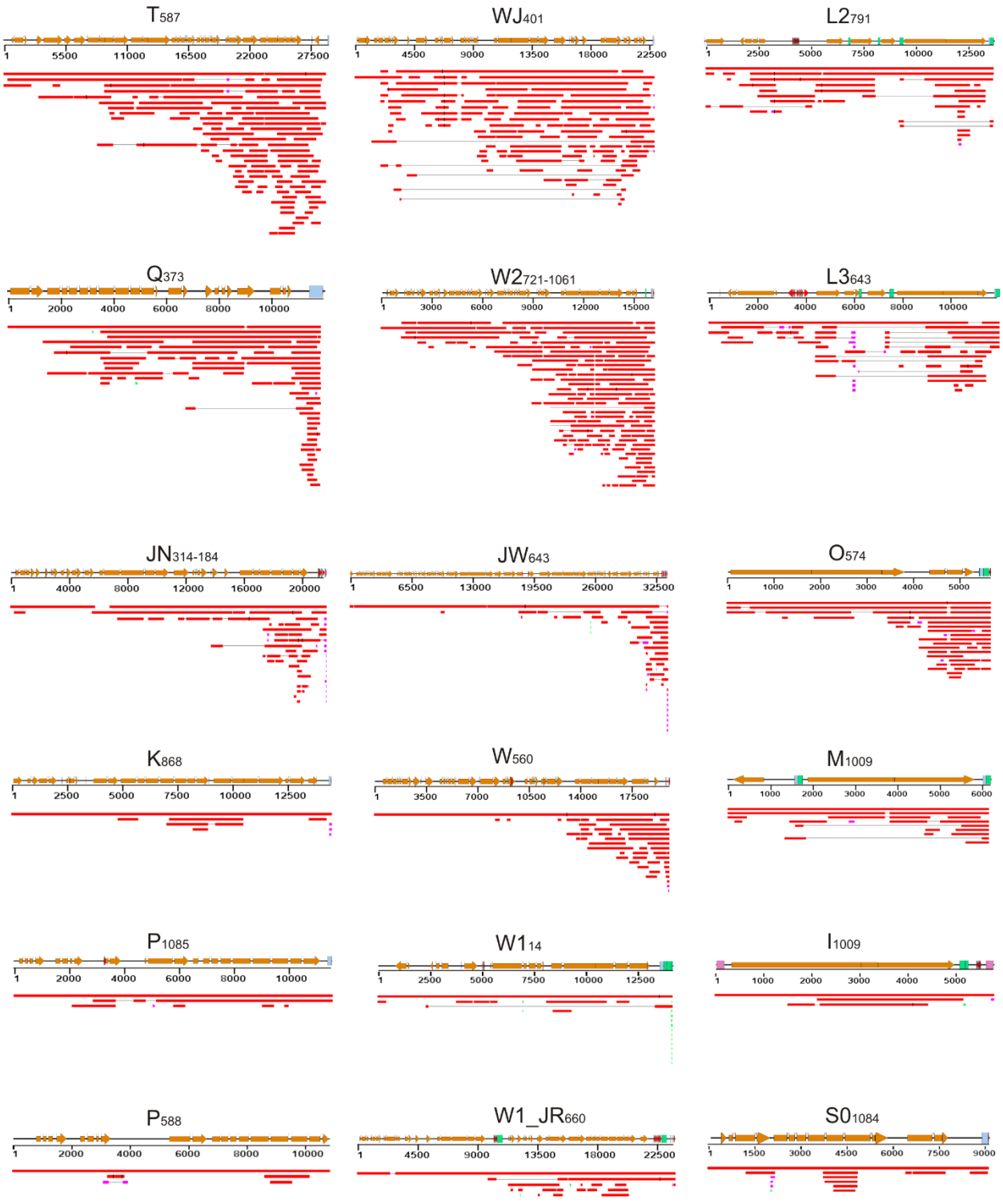
T587
1 5500 11000 16500 22000 27500
WJ401
1 4500 9000 13500 18000 22500
L2791
1 2500 5000 7500 10000 12500
Q373
1 2000 4000 6000 8000 10000
W2721-1061
1 3000 6000 9000 12000 15000
L3643
1 2000 4000 6000 8000 10000
JN314-184
1 4000 8000 12000 16000 20000
JW643
1 6500 13000 19500 26000 32500
O574
1 1000 2000 3000 4000 5000
K868
1 2500 5000 7500 10000 12500
W560
1 3500 7000 10500 14000 17500
M1009
1 1000 2000 3000 4000 5000 6000
P1085
1 2000 4000 6000 8000 10000
W114
1 2500 5000 7500 10000 12500
I1009
1 1000 2000 3000 4000 5000
P588
1 2000 4000 6000 8000 10000
W1_JR660
1 4500 9000 13500 18000 22500
S01084
1 1500 3000 4500 6000 7500 9000

Fig. S2

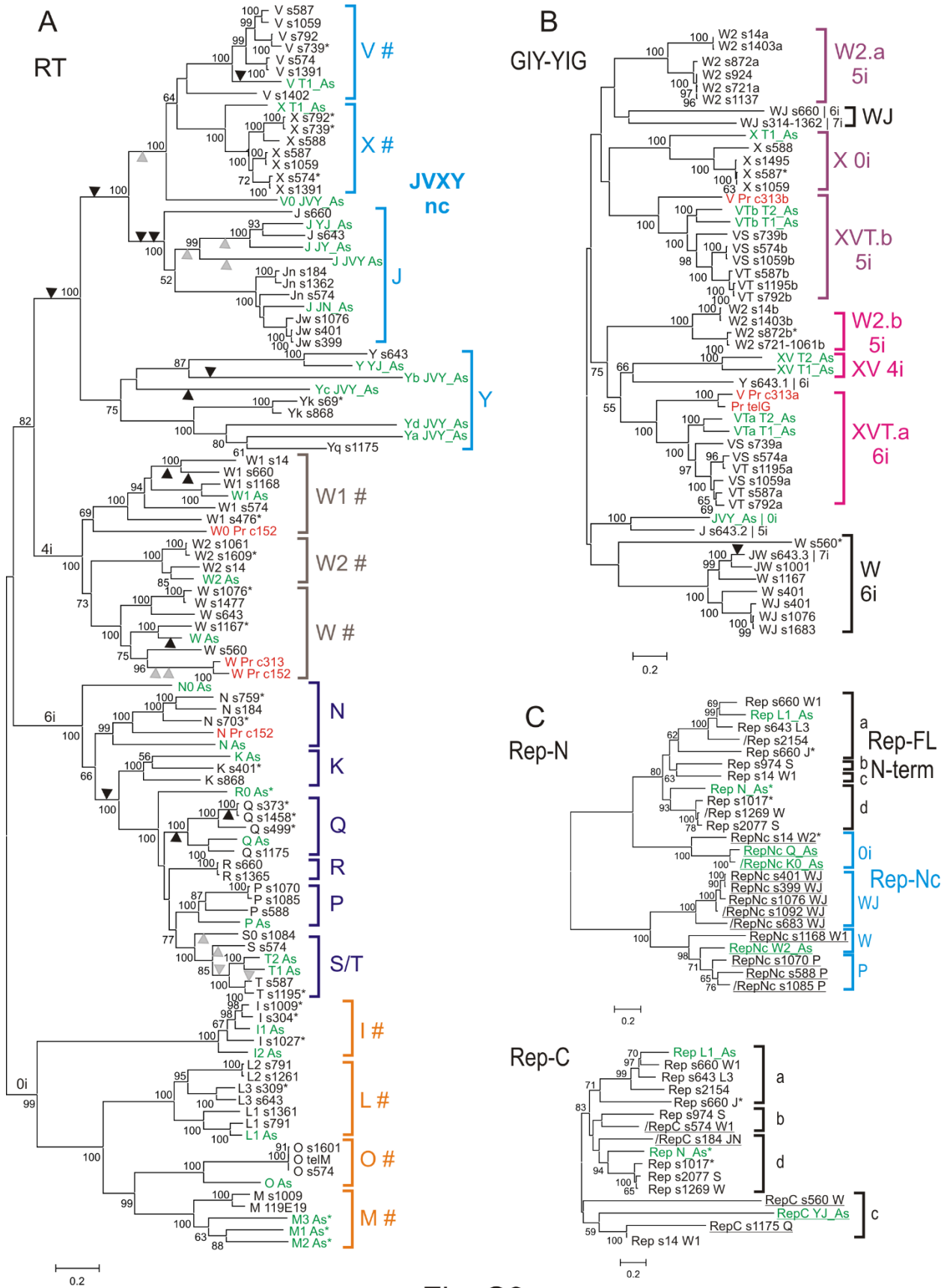
A
RT
V #
X #
JVXY
nc
J
Y
W1 #
W2 #
W #
N
K
Q
R
P
S/T
I #
L #
O #
M #
4i
6i
0i
0.2
B
GIY-YIG
W2.a
5i
WJ
X 0i
XVT.b
5i
W2.b
5i
XV 4i
XVT.a
6i
W
6i
0.2
C
Rep-N
a
b
c
d
Rep-FL
N-term
0i
Rep-Nc
WJ
W
P
0.2
Rep-C
a
b
d
c
0.2

Fig. S3

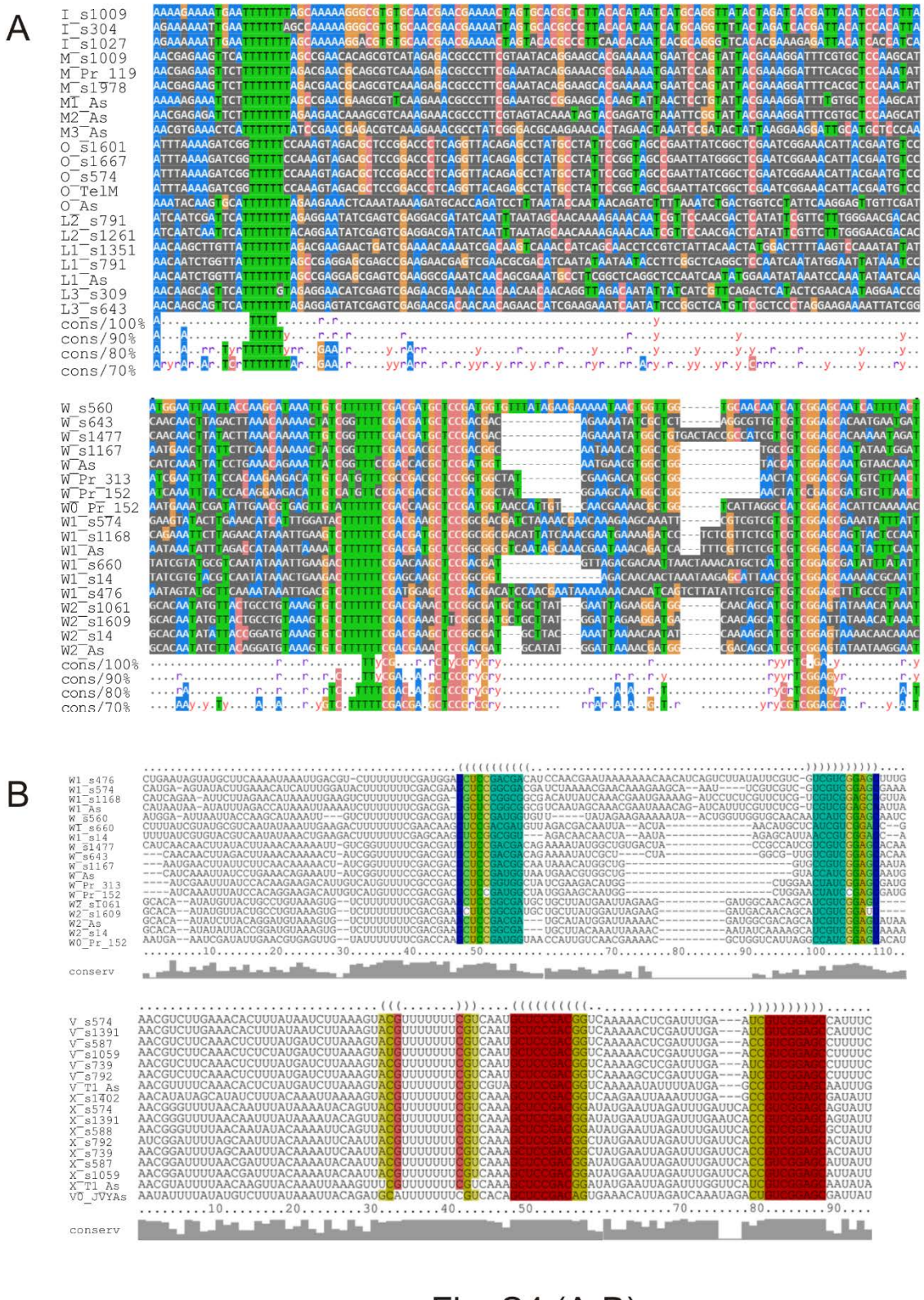
A
I_s1009
I_s304
I_s1027
M_s1009
M_Pr_119
M_s1978
MI_As
M2_As
M3_As
O_s1601
O_s1667
O_s574
O_TelM
O_As
L2_s791
L2_s1261
L1_s1351
L1_s791
L1_As
L3_s309
L3_s643
cons/100%
cons/90%
cons/80%
cons/70%
W_s560
W_s643
W_s1477
W_s1167
W_As
W_Pr_313
W_Pr_152
W0_Pr_152
W1_s574
W1_s1168
W1_As
W1_s660
W1_s14
W1_s476
W2_s1061
W2_s1609
W2_s14
W2_As
cons/100%
cons/90%
cons/80%
cons/70%
B
W1_s476
W1_s574
W1_s1168
W1_As
W_s560
W1_s660
W1_s14
W_s1477
W_s643
W_s1167
W_As
W_Pr_313
W_Pr_152
W2_s1061
W2_s1609
W2_As
W2_s14
W0_Pr_152
conserv
V_s574
V_s1391
V_s587
V_s1059
V_s739
V_s792
V_T1_As
X_s1402
X_s574
X_s1391
X_s588
X_s792
X_s739
X_s587
X_s1059
X_T1_As
V0_JVYAs
conserv

Fig. S4 (A-B)

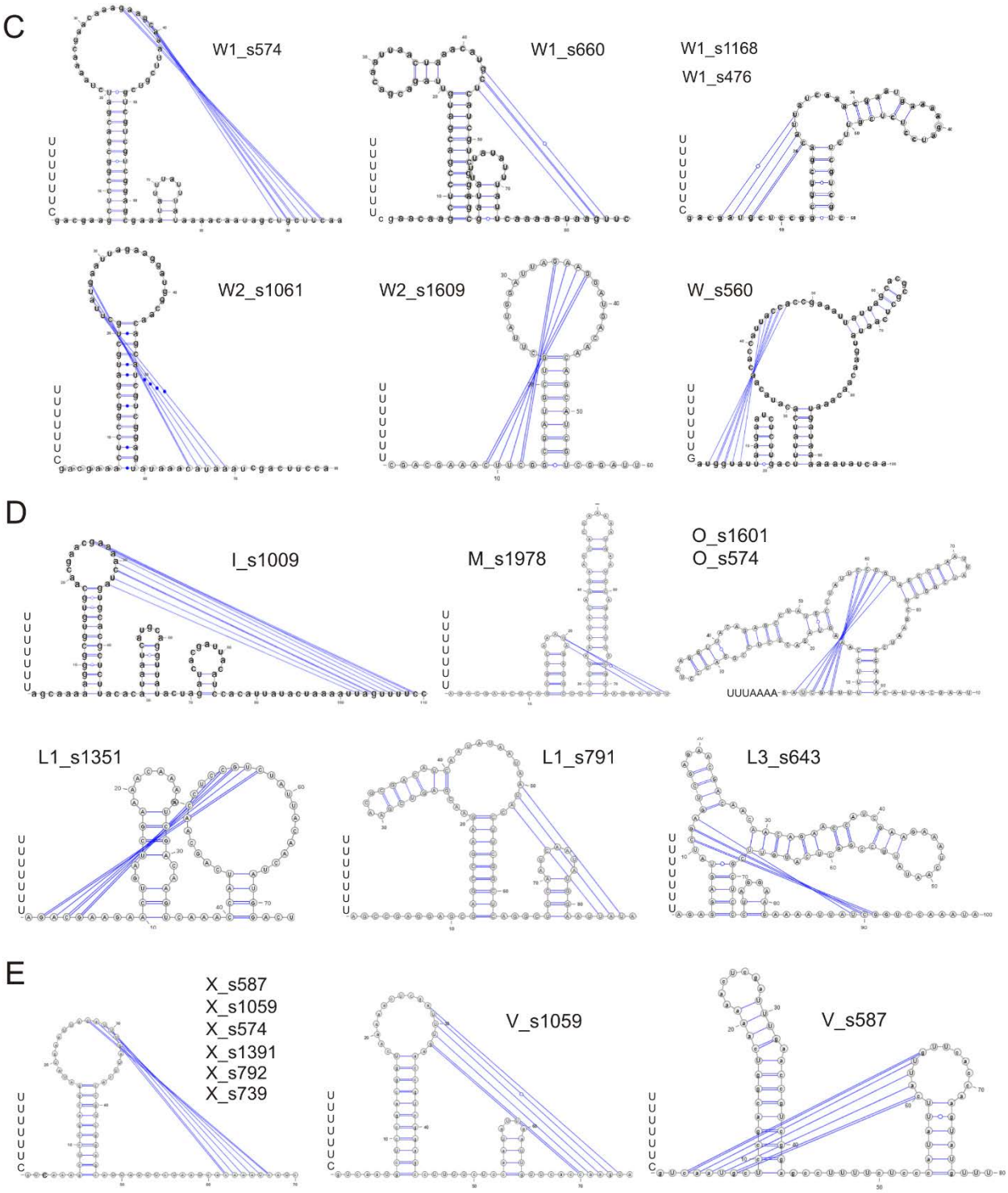
C
W1_s574
W1_s660
W1_s1168
W1_s476
W2_s1061
W2_s1609
W_s560
D
I_s1009
M_s1978
O_s1601
O_s574
L1_s1351
L1_s791
L3_s643
E
X_s587
X_s1059
X_s574
X_s1391
X_s792
X_s739
V_s1059
V_s587

Fig. S4 (C-E)

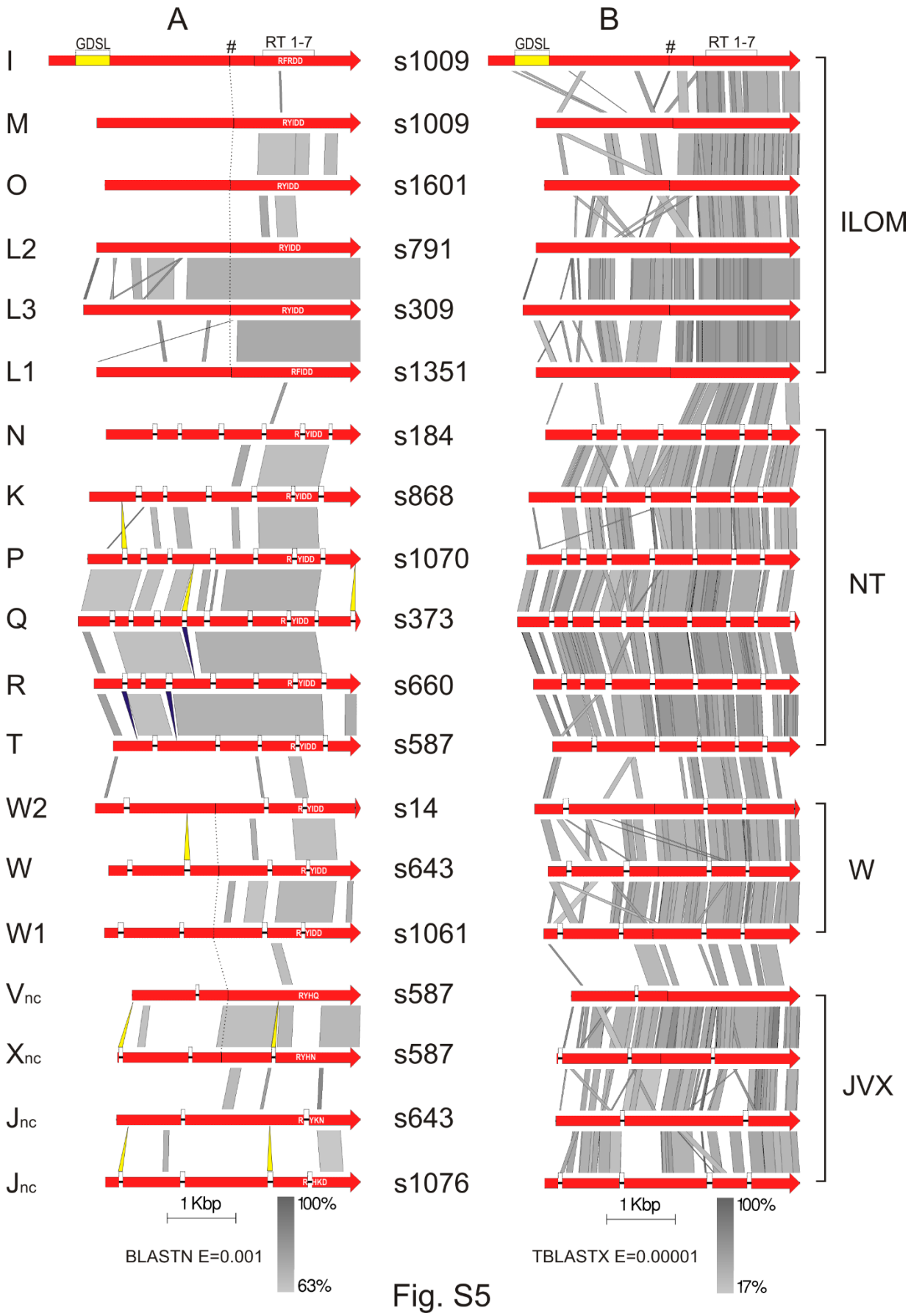
A
B
GDSL
#
RT 1-7
I
M
O
L2
L3
L1
N
K
P
Q
R
T
W2
W
W1
Vnc
Xnc
Jnc
Jnc
RFRDD
RYIDD
RFIDD
RYHQ
RYHN
RYKN
RHKD
s1009
s1009
s1601
s791
s309
s1351
s184
s868
s1070
s373
s660
s587
s14
s643
s1061
s587
s587
s643
s1076
ILOM
NT
W
JVX
1 Kbp
100%
63%
BLASTN E=0.001
1 Kbp
100%
17%
TBLASTX E=0.00001

Fig. S5

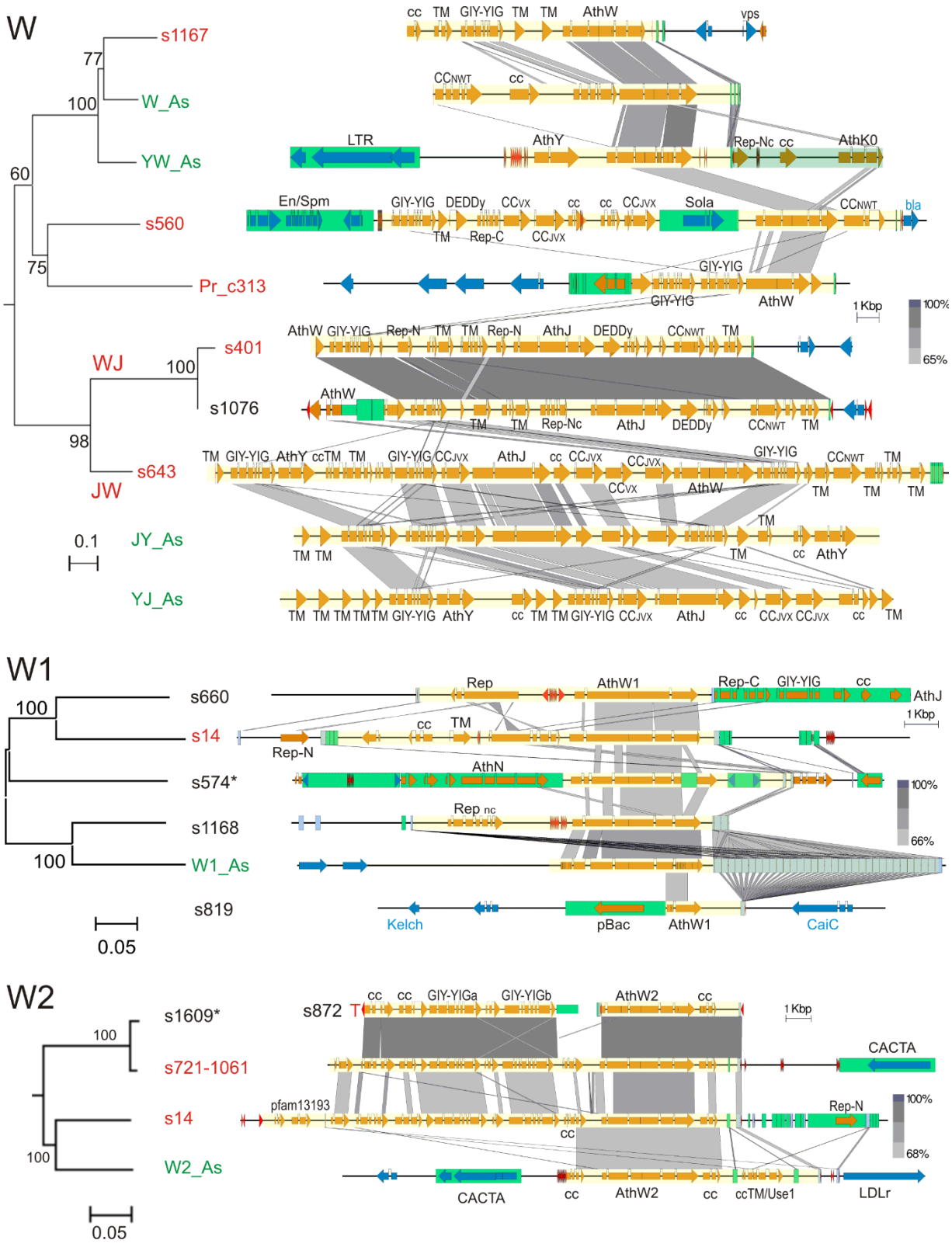
W
s1167
77
W_As
100
YW_As
60
s560
75
Pr_c313
WJ
100
s401
s1076
98
s643
JW
0.1
JY_As
YJ_As
W1
100
s660
s14
s574*
s1168
100
W1_As
s819
0.05
W2
s1609*
100
s721-1061
s14
100
W2_As
0.05
1 Kbp
100%
65%
66%
68%
LTR
AthY
Rep-Nc
AthK0
En/Spm
Sola
bla
Rep-C
AthW
AthJ
DEDDy
Rep-N
Rep
AthW1
Rep-C
AthN
Kelch
pBac
CaiC
s872
CACTA
pfam13193
LDLr

Fig. S6 (A)

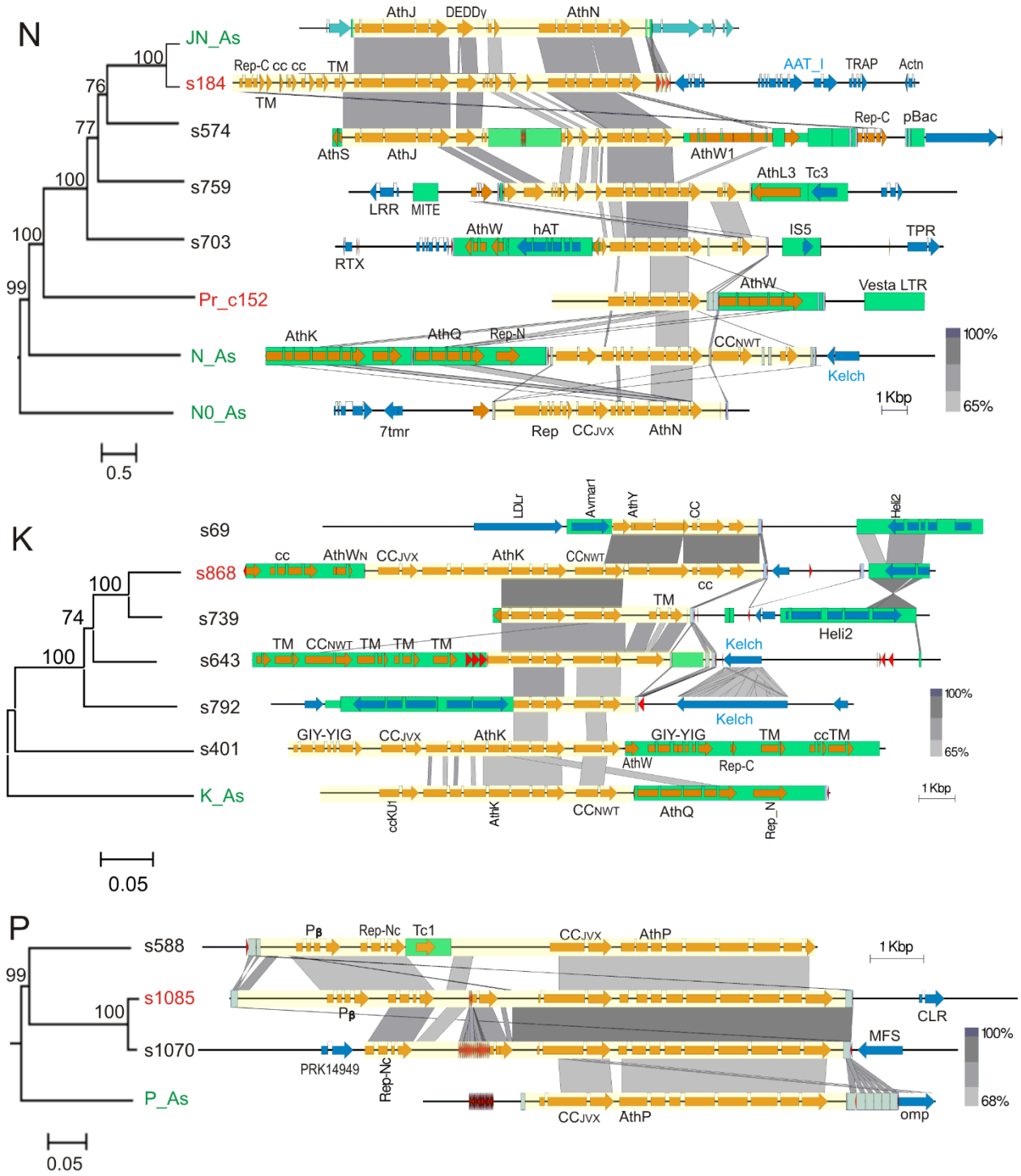
N
JN_As
s184
s574
s759
s703
Pr_c152
N_As
N0_As
100
76
77
100
100
99
0.5
AthJ
DEDDy
AthN
Rep-C cc cc
TM
TM
AAT_I
TRAP
Actn
Rep-C
pBac
AthS
AthJ
AthW1
AthL3
Tc3
LRR
MITE
AthW
hAT
IS5
TPR
RTX
AthW
Vesta LTR
AthK
AthQ
Rep-N
CCNWT
Kelch
100%
65%
1 Kbp
7tmr
Rep
CCJVX
AthN
K
s69
s868
s739
s643
s792
s401
K_As
100
74
100
0.05
LDLr
Avrmar1
AthY
CC
Heli2
cc
AthWN
CCJVX
AthK
CCNWT
cc
TM
Heli2
TM
CCNWT
TM
TM
TM
Kelch
Kelch
100%
65%
GIY-YIG
CCJVX
AthK
GIY-YIG
TM
ccTM
AthW
Rep-C
1 Kbp
ccKU
AthK
CCNWT
AthQ
Rep_N
P
s588
s1085
s1070
P_As
99
100
0.05
Pβ
Rep-Nc
Tc1
CCJVX
AthP
1 Kbp
Pβ
CLR
MFS
100%
68%
PRK14949
Rep-Nc
CCJVX
AthP
omp

Fig. S6 (B)

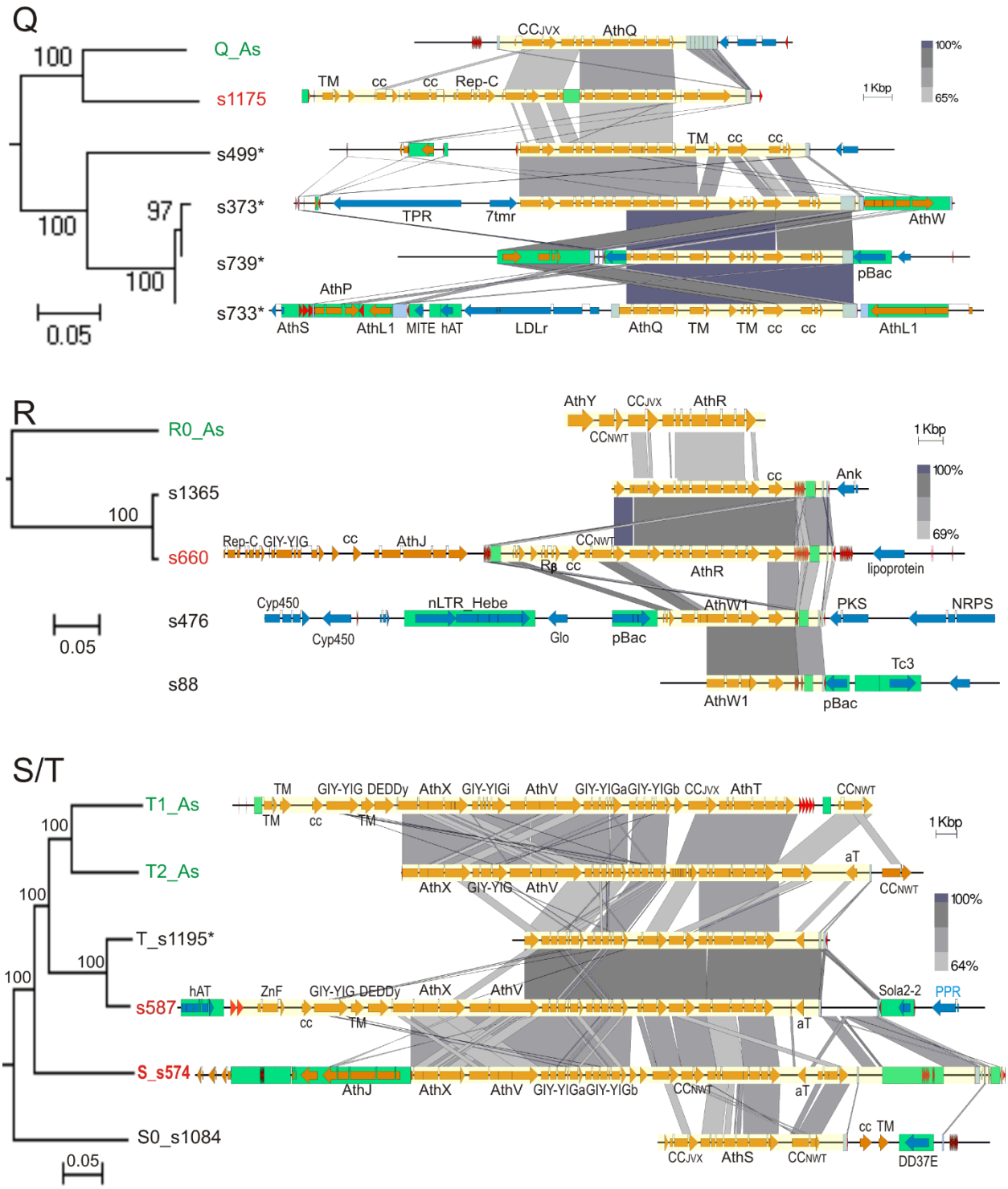
Q
100
Q_As
s1175
s499*
97
100
s373*
100
s739*
s733*
0.05
CCJVX
AthQ
TM
cc
cc
Rep-C
1 Kbp
100%
65%
TM
cc
cc
TPR
7tmr
AthW
pBac
AthP
AthS
AthL1
MITE
hAT
LDLr
AthQ
TM
TM
cc
cc
AthL1
R
R0_As
s1365
100
s660
s476
0.05
s88
AthY
CCJVX
AthR
CCNWT
1 Kbp
cc
Ank
100%
69%
Rep-C
GIY-YIG
cc
AthJ
CCNWT
Rβ
cc
AthR
lipoprotein
Cyp450
nLTR
Hebe
AthW1
PKS
NRPS
Cyp450
Glo
pBac
Tc3
AthW1
pBac
S/T
T1_As
100
T2_As
100
T_s1195*
100
100
s587
S_s574
S0_s1084
0.05
TM
GIY-YIG
DEDDy
AthX
GIY-YIGi
AthV
GIY-YIGa
GIY-YIGb
CCJVX
AthT
CCNWT
TM
cc
TM
1 Kbp
aT
AthX
GIY-YIG
AthV
CCNWT
100%
64%
hAT
ZnF
GIY-YIG
DEDDy
AthX
AthV
Sola2-2
PPR
cc
TM
aT
AthJ
AthX
AthV
GIY-YIGa
GIY-YIGb
CCNWT
aT
cc
TM
CCJVX
AthS
CCNWT
DD37E

Fig. S6 (C)

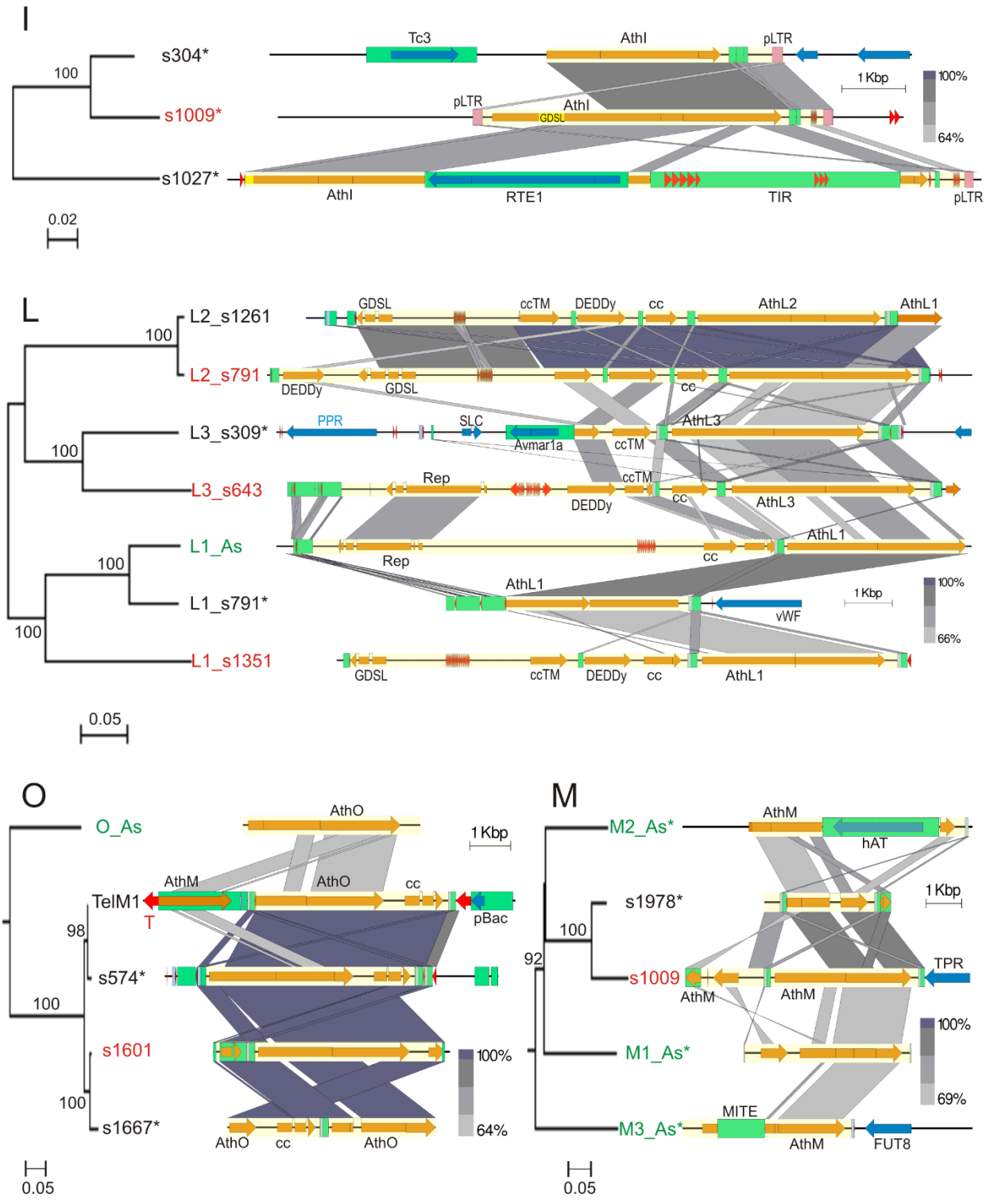
I
s304*
s1009*
s1027*
100
Tc3
AthI
pLTR
1 Kbp
100%
64%
GDSL
RTE1
TIR
0.02
L
L2_s1261
L2_s791
L3_s309*
L3_s643
L1_As
L1_s791*
L1_s1351
GDSL
ccTM
DEDDy
cc
AthL2
AthL1
PPR
SLC
Avmar1a
AthL3
Rep
vWF
66%
0.05
O
O_As
TelM1
s574*
s1601
s1667*
AthO
AthM
pBac
98
M
M2_As*
s1978*
s1009
M1_As*
M3_As*
hAT
TPR
MITE
FUT8
69%
92

Fig. S6 (D)

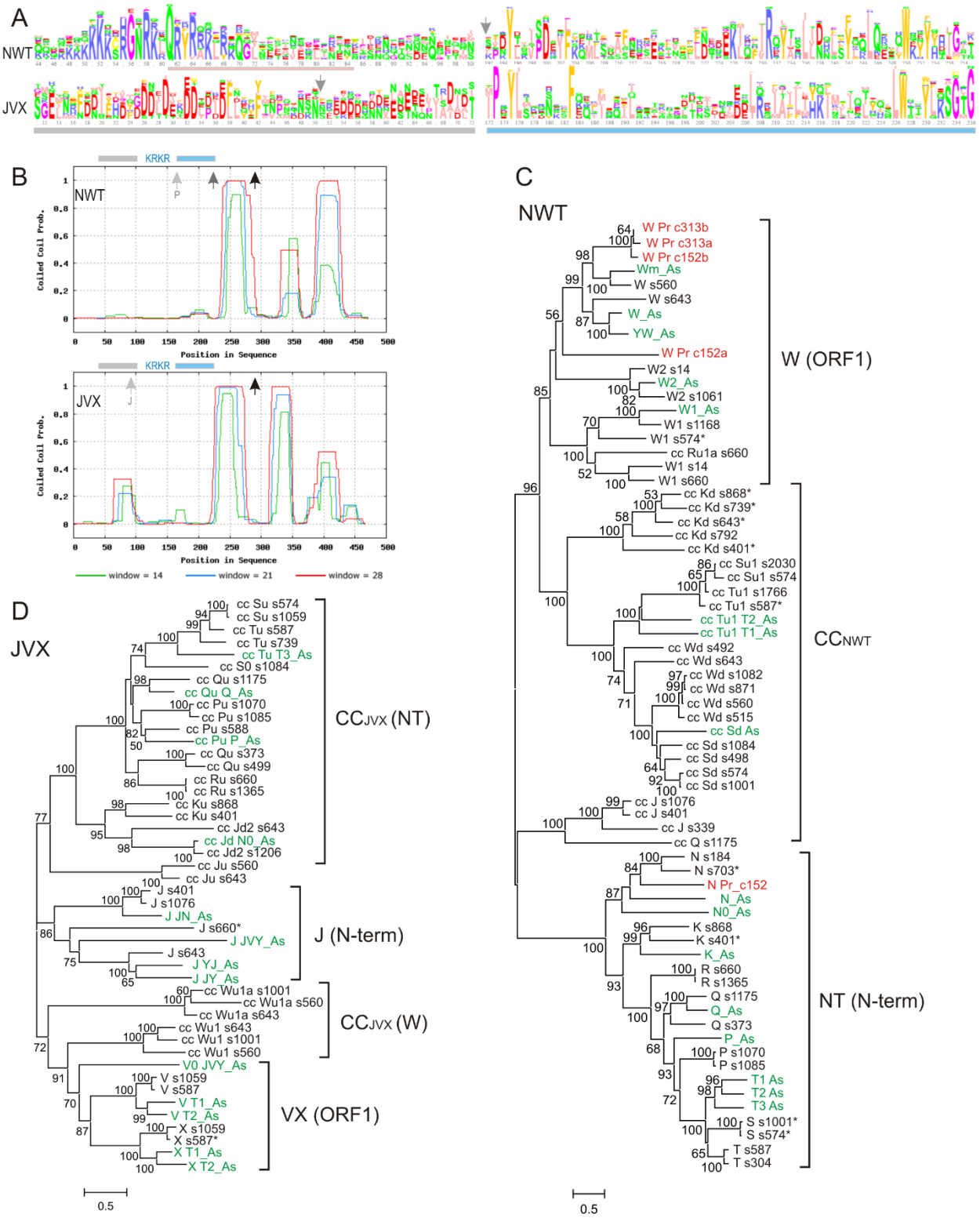
A
NWT
JVX
B
NWT
JVX
KRKR
Coiled Coil Prob.
Position in Sequence
window = 14
window = 21
window = 28
C
NWT
W (ORF1)
CCNWT
NT (N-term)
W Pr c313b
W Pr c313a
W Pr c152b
Wm_As
W s560
W s643
W_As
YW_As
W Pr c152a
W2 s14
W2_As
W2 s1061
W1_As
W1 s1168
W1 s574*
cc Ru1a s660
W1 s14
W1 s660
cc Kd s868*
cc Kd s739*
cc Kd s643*
cc Kd s792
cc Kd s401*
cc Su1 s2030
cc Su1 s574
cc Tu1 s1766
cc Tu1 s587*
cc Tu1 T2_As
cc Tu1 T1_As
cc Wd s492
cc Wd s643
cc Wd s1082
cc Wd s871
cc Wd s560
cc Wd s515
cc Sd As
cc Sd s1084
cc Sd s498
cc Sd s574
cc Sd s1001
cc J s1076
cc J s401
cc J s339
cc Q s1175
N s184
N s703*
N Pr_c152
N_As
N0_As
K s868
K s401*
K_As
R s660
R s1365
Q s1175
Q_As
Q s373
P_As
P s1070
P s1085
T1 As
T2 As
T3 As
S s1001*
S s574*
T s587
T s304
0.5
D
JVX
CCJVX (NT)
J (N-term)
CCJVX (W)
VX (ORF1)
cc Su s574
cc Su s1059
cc Tu s587
cc Tu s739
cc Tu T3_As
cc S0 s1084
cc Qu s1175
cc Qu Q_As
cc Pu s1070
cc Pu s1085
cc Pu s588
cc Pu P_As
cc Qu s373
cc Qu s499
cc Ru s660
cc Ru s1365
cc Ku s868
cc Ku s401
cc Jd2 s643
cc Jd N0_As
cc Jd2 s1206
cc Ju s560
cc Ju s643
J s401
J s1076
J JN_As
J s660*
J JVY_As
J s643
J YJ_As
J JY_As
cc Wu1a s1001
cc Wu1a s560
cc Wu1a s643
cc Wu1 s643
cc Wu1 s1001
cc Wu1 s560
V0 JVY_As
V s1059
V s587
V T1_As
V T2_As
X s1059
X s587*
X T1_As
X T2_As
0.5

Fig. S7

Table S1. Gene content at informative 3'-junctions of *Terminon* elements (within 10 kb of the 3'end). The remaining uninformative 3'-junctions contain little or no adjacent flanking DNA and are not shown. TR, telomeric repeats at 3'-ends. Color coding, genes of foreign origin (blue, bacterial; purple, fungal; green, plant; pink, protist) or TEs (yellow, PLE; orange, DNA TE). Highlighted copies are located on allelic scaffolds. +/-, sense or antisense CDS orientation; /, truncation.

Copy	Scaffold	TR	3'-CDS	+/-	Copy	Scaffold	TR	3'-CDS	+/-
JW	643	✔	/Athena-K	-	W	373	✔	TPR	-
JW	309	✔	/TPR	+	W	373	✔	SET	-
JW	693	✔	NAD-ADPRT/	-	W	1354	✔	NRPS	-
JW	574	✔	UDP-GTase	+	W	922	✔	/LDL receptor	-
WJ	664	✔	lectin	+	W	560	✔	/β-lactamase	+
WJ	399	✔	Kelch	-	W	476	✔	NRPS	-
WJ	1070	✔	TPR	-	W	1082	✔	RfbB epimerase	-
WJ	1571	✔	NRPS	-	W	664	✔	ABC transporter	-
WJ	1292	✔	TPR	-	W2	1041	✔	LRR	-
WJ	1203	✔	ankyrin	-	W1	88	✔	Pokey4 TPase/	-
WJ	1386	✔	hypothetical Av	+	W2	536	✔	GTase	-
WJ	1058	✔	GTase	-	W2	339	✔	vcbs/SxtP	-
WJ	3004	✔	AIG1	-	W2	552	✔	zf-C2H2/	-
WJ	860	✔	MTase	+	W2	214	✔	Ig-like	-
WJ	648	✔	GTPase	-	W2	274	✔	Ig-like	-
JN	184	✔	lectin	-	S	852	✔	NRPS	+
N	695	✔	NRPS	-	S	1254	✔	NRPS	+
N	703	✔	IS5	-	S	1001	✔	/Athena-W	+
N	759	✔	Athena-M/	-	S	498	-	vcbs	-
N,W	736	✔	GTase	-	S	974	✔	LDL receptor	-
K	868	✔	hypothetical	-	T	1084	✔	hypothetical	-
K	643	✔	Kelch/	-	T	792	✔	PAT1	-
K	792	✔	Kelch	-	T	209	✔	ABC ATPase	-
P	1070	✔	folate transport	-	T	404	✔	/Athena-K	-
P	1085	✔	lectin receptor	+	T	979	✔	acyltransferase	-
Q	739	✔	Looper TPase/	-	T	587	✔	Sola2-2	-
Q	88	✔	Merlin TPase	-	T	809	✔	vcbs	-
Q	339	✔	hypothetical	-	T	41	-	/filamin	+
Q	156	✔	Merlin TPase	-	T	1041	✔	F-box/LRR	-
R	660	✔	lipoprotein	-	T	304	✔	Penelope2	-
R	1365	✔	ankyrin	-	T	974	✔	LDL receptor	-
L1	560	✔	/Helitron	+	T	1457	✔	Athena-L	+
L1	791	-	cell surface prot	-	T	612	✔	FkbM MTase	+
L1	733	+/-	Athena-Q	-	T	1027	✔	NHL	+
L1	404	-	Avmar TPase	+	T	798	✔	nitroreductase	-
L2	866	✔	peptidase	-	M	1009	-	TPR/	-
L2	24	✔	/NHL	+	M	1224	-	MuDR TPase	-
L2	1015	-	hypothetical Av	+	M	1036	-	TLR-2	-
L2	1193	-	NRPS	-	M	809	-	ISL2EU TPase	+
L2	476	✔	7tm receptor	+	M	13	-	NAD-ADPRT	+
L2	1261	-	/Athena-L	+	M	1978	✔	Athena-M	+
L3	309	✔	hypothetical Av	-	O	574	✔	hAT TPase	+
L3	1045	✔	Sola2d		O	399	✔	Athena-J	+
L3	643	✔	Athena-JW		O	404	✔	Tcb1 TPase	+
L3	664	✔	peptidase		O	703	✔	Athena-L	+

Table S2. Differences in telomeric repeat distribution around different TE types.

Window size	2Kb		5Kb		10Kb	
ANOVA Tukey's comparisons	t value	p value	t value	p value	t value	p value
LTR – Helitron	-1.357	0.731	-1.042	0.89209	-1.687	0.5125
non-LTR – Helitron	-1.268	0.783	-1.588	0.57897	-1.135	0.8523
Athena – Helitron	4.01	<0.001 ***	5.88	< 0.001 ***	5.678	<0.001 ***
Penelope – Helitron	1.244	0.796	-0.021	1	0.163	1
TIR – Helitron	-0.679	0.982	-0.298	0.99964	0.925	0.9326
non-LTR – LTR	0.12	1	-0.453	0.99723	0.541	0.9936
Athena – LTR	4.804	<0.001 ***	6.173	< 0.001 ***	6.587	<0.001 ***
Penelope – LTR	2	0.317	0.605	0.98929	1.167	0.8368
TIR – LTR	1.076	0.879	1.006	0.90585	2.753	0.0582 .
Athena – non-LTR	4.819	<0.001 ***	6.819	< 0.001 ***	6.219	<0.001 ***
Penelope – non-LTR	1.945	0.349	0.912	0.93636	0.823	0.9584
TIR – non-LTR	0.966	0.92	1.675	0.52085	2.152	0.2394
Penelope – Athena	-1.144	0.848	-3.462	0.00619 **	-3.166	0.0170 *
TIR – Athena	-5.442	<0.001 ***	-7.391	< 0.001 ***	-6.111	<0.001 ***
TIR – Penelope	-1.63	0.551	-0.111	1	0.242	0.9999

A one-way ANOVA followed by Tukey's post hoc test was used to test for differences in the mean counts of telomeric repeats between different TE types (Heliron, LTR, non-LTR, Athena, Penelope and TIR) in three window sizes. TE type was found to have an impact on the number of nearby telomeric repeats: 2 Kb window, $F(5, 15847) = 7.366$, ***$P < 0.001$; 5 Kb window, $F(5, 15847) = 12.788$, ***$P < 0.001$; and 10 Kb window, $F(5, 15847) = 11.461$, ***$P < 0.001$. Post-hoc Tukey's test showed that Athena RTs have significantly higher counts of nearby telomeric repeats than LTR, non-LTR, TIR and Helitron elements, while differences with Penelope elements start at the 5 Kb window. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values are reported.

Supplementary Methods

Genome and transcriptome datasets. We used the *Adineta vaga* reference genome (Flot, et al. 2013) available in GenBank as unannotated scaffolds (CAWI000000000.2), for homology searches, and in the browser format at www.genoscope.cns.fr/adineta/cgi-bin/gbrowse/adineta to download the annotated scaffolds containing matches to *Athena* RTs (Gladyshev and Arkhipova 2007). Matching scaffolds were individually reannotated using GeneWise (www.ebi.ac.uk/Tools/psa/genewise/) and manually adjusted to introduce corrections into *Athena* RTs and associated ORF sequences, which were mostly misannotated due to the poor approximation of the *C. elegans*-trained gene models to TE-encoded ORFs. For the same reason, this procedure was also applied to genes of non-metazoan origin. The Sanger-sequenced *A. vaga* and *P. roseola* large-insert library clones were from Av_184A11, Av_119E19, Pr_152C9, Pr_313N6, PrTEL_IV_4, Pr_TEL_G (EU643486, EF485018, DQ138288, MF143428, EF485015, EF485006). The PacBio draft partial assembly was obtained from 15 SMRT cells using the HGAP assembler in the SMRT® Portal (Pacific Biosciences). TE consensus sequences reported in this study have been submitted to Repbase (Bao, et al. 2015). The *A. vaga* transcripts were uniquely mapped to the reference genome using Bowtie2 (Langmead and Salzberg 2012) on 75-nt Illumina GAIIx reads for cDNA (ERX234948), and using Bowtie (Langmead, et al. 2009) on 50-nt Illumina HiSeq small RNA reads (SRP070765). Aligned sequence reads were counted by annotated genomic feature with htseq-count (Anders, et al. 2015).

Bioinformatics. Programmed ribosomal frameshifting sites with simple pseudoknots were identified using KnotInFrame (Janssen and Giegerich 2015) and graphically represented using VARNA (Darty, et al. 2009). Secondary-structure-based multiple RNA alignments were performed with LocARNA (Smith, et al. 2010). Domain architecture was assessed with CDART (Geer, et al. 2002). Graphical representation of pairwise BLASTN and TBLASTX similarities between genomic scaffolds was done with Easyfig 2.2.2 (Sullivan, et al. 2011), and graphical overviews of genome-wide BLASTN searches of *A. vaga* scaffolds were obtained from NCBI server (https://blast.ncbi.nlm.nih.gov/Blast.cgi) with the megablast option. Multiple sequence alignments were done using MUSCLE (Edgar 2004); for protein-coding sequences, alignments were codon-based. Sequence logos for multiple alignments with AlignmentViewer, prediction of coiled-coil motifs with COILS/PCOILS, and profile-profile homology searches with HHPRED were done using the MPI toolkit (Alva, et al. 2016). HHR motif alignment was created with Boxshade 3.3.1. Maximum likelihood phylogenetic analysis was done using the GTR+F+G model for nucleotide sequences and the WAG+F+G model for amino acid sequences, with a discrete gamma distribution used to model evolutionary rate differences across sites (G=5), and the resulting trees were edited in MEGA 7.0.18 (Kumar, et al. 2016). To identify HHR motifs characteristic of the L and WJ families, we used RNAMotif 3.1.1 (Macke, et al. 2001) with the descriptor from (Cervera and De la Pena 2014) relaxed to accommodate core-I: ss (minlen=7, maxlen=9); longer loop 2: ss (minlen=3, maxlen=16); and split core-II: ss (minlen=3, maxlen=6, seq="gaa$"), ss (minlen=3, maxlen=10, seq="nuh$"). Telomeric repeats in *A. vaga* were annotated using the sequence $(TGWGGG)_n$ and counted in different window sizes around the annotated genomic features of interest. The *A. vaga* gene set was divided into non-metazoan (foreign) and metazoan subsets, with the latter additionally subdivided to genes with and without piRNA coverage, as described in (Rodriguez and Arkhipova 2016). Statistical significance of the frequency of association between telomeric repeats and different TE types was estimated using single-factor ANOVA (Tukey's test). Significant values were assumed at $p < 0.05$.

Nucleic acid manipulations. Genomic DNA from rotifer eggs purified as in (Flot, et al. 2013) and ground in liquid N_2 was extracted with the Qiagen Genomic-Tip 500/G tissue protocol according to manufacturer's instructions with minor modifications, and checked for integrity by pulsed-field gel electrophoresis. After BluePippin size selection, it was used to construct a 20-kb library, which was sequenced on PacBio RS II at the Johns Hopkins University Deep Sequencing and Microarray Core with P6-C4 chemistry. The 462-bp JW-643 fragment with three identical tandem HHR-containing units from scaffold_643:36798..37259 was chemically synthesized by GenScript and cloned into the KpnI-BamHI sites of pBluescript II SK+. The plasmid template was linearized with BamHI and used for *in vitro* transcription. The reaction mix containing 40 mM Tris–HCl, pH 8.0, 6 mM $MgCl_2$, 2 mM spermidine, 0.5 mg/ml RNase-free BSA, 0.1% Triton X-100, 10 mM dithiothreitol, 1 mM ATP, 1 mM CTP, 1 mM GTP, 0.1 mM UTP, 0.5 µCi/µl [α-^{32}P]-UTP (Perkin-Elmer), 20 U RNase inhibitor (Roche), 20 µg/ml DNA, and 4 U/µl T7 RNA polymerase (Stratagene) was incubated at 37°C for 8.5 hrs. Labeled full-length transcripts were extracted from polyacrylamide gels and purified using RNA Clean & Concentrator™-5 kit (Zymo Research). Purified *in*

vitro transcripts were used to set up self-cleavage reactions in 50 mM Tris-HCl, pH 7.5, 20 U RNase inhibitor (Roche), heated at 95°C for 1 min, and pre-incubated at 25°C for 15 min. Self-cleavage was initiated by addition of $MgCl_2$ or $MnCl_2$, aliquots were taken at different time points, immediately placed on ice, mixed with the stop solution (8 M urea, 50% formamide, 50 mM EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue), and run on 15% urea PAGE. Self-cleavage assays for *AvPen3a* HHR were essentially similar, except that the plasmid template was obtained by cloning of a 310-bp PCR-generated *AvPen3a* (scaffold 1009) using primers F (CGGGGTACCTGCAGTAAAAACAAAACAGAATGAAT) and R2 (CGCGGATCCTGCAGACGGTCCTTGATGTT), and self-cleavage was monitored by extension of the R2 primer end-labeled with [γ-^{32}P]-ATP (Perkin-Elmer) and T4 polynucleotide kinase (NEB).

Supplementary References

Alva V, Nam S-Z, Söding J, Lupas AN. 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res 44:W410-W415.

Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31:166-169.

Arkhipova I. 2006. Distribution and phylogeny of Penelope-like elements in eukaryotes. Syst Biol 55:875-885.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:11.

Cervera A, De la Pena M. 2014. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. Mol Biol Evol 31:2941-2947.

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics 25:1974-1975.

Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 5:113.

Flot JF, Hespeels B, Li X, Noel B, Arkhipova I, Danchin EG, Hejnol A, Henrissat B, Koszul R, Aury JM, et al. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature 500:453-457.

Geer LY, Domrachev M, Lipman DJ, Bryant SH. 2002. CDART: protein homology by domain architecture. Genome Res 12:1619-1623.

Gladyshev E, Arkhipova IR. 2007. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. Proc Natl Acad Sci U S A 104:9352-9357.

Janssen S, Giegerich R. 2015. The RNA shapes studio. Bioinformatics 31:423-425.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870-1874.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Meth 9:357-359.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10.

Macke TJ, Ecker DJ, Gutell RR, Gautheret D, Case DA, Sampath R. 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Research 29:4724-4735.

Rodriguez F, Arkhipova IR. 2016. Multitasking of the piRNA silencing machinery: Targeting transposable elements and foreign genes in the bdelloid rotifer Adineta vaga. Genetics 203:255-268.

Smith C, Heyne S, Richter AS, Will S, Backofen R. 2010. Freiburg RNA Tools: a web server integrating IntaRNA, ExpaRNA and LocARNA. Nucleic Acids Res 38:W373-W377.

Sullivan MJ, Petty NK, Beatson SA. 2011. Easyfig: a genome comparison visualizer. Bioinformatics 27:1009-1010.

Xiong Y, Eickbush TH. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. The EMBO Journal 9:3353-3362.