

# Supplementary Figure S1

A

Slyc = *S. lycopersicum* Paraseq  
Spen = *S. pennellii* Paraseq  
Stub = *S. tuberosum* Paraseq  
Smel = *S. melongena* Paraseq

*Florendovirus* C1+C2

*Caulimovirus*/  
*Soymovirus*-related

*Solendovirus* C2

*Solendovirus* C1

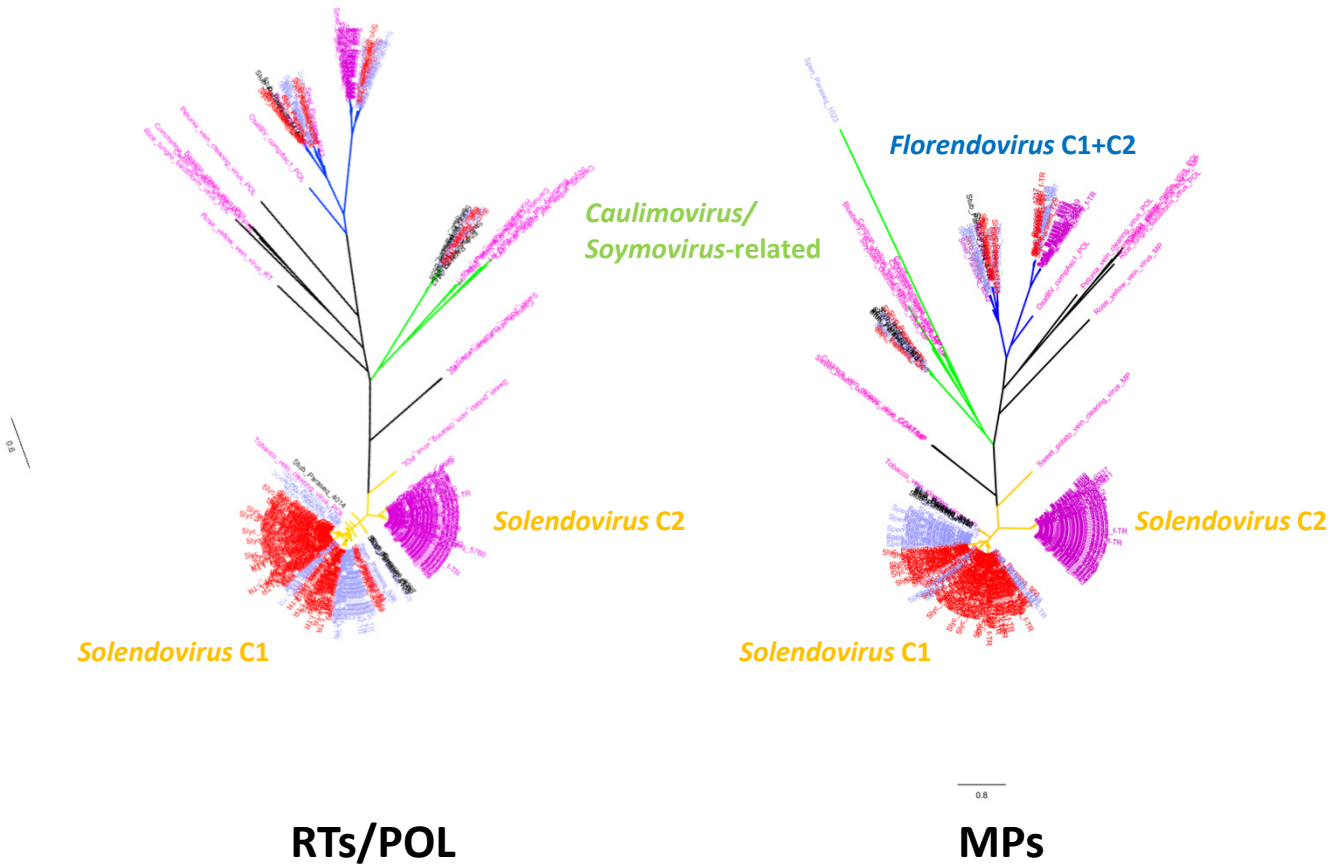
RTs/POL

*Florendovirus* C1+C2

*Solendovirus* C2

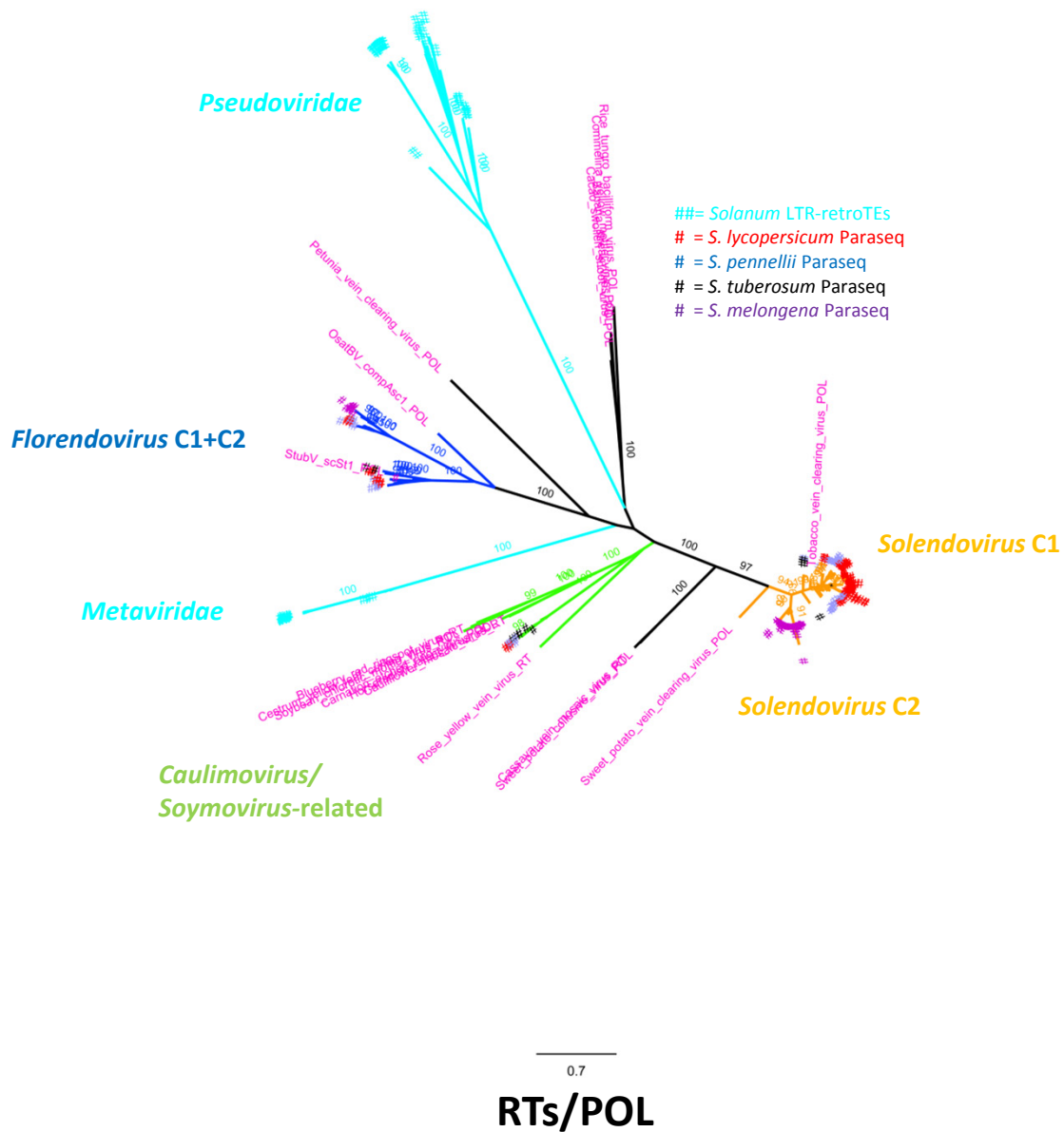
*Solendovirus* C1

MPs



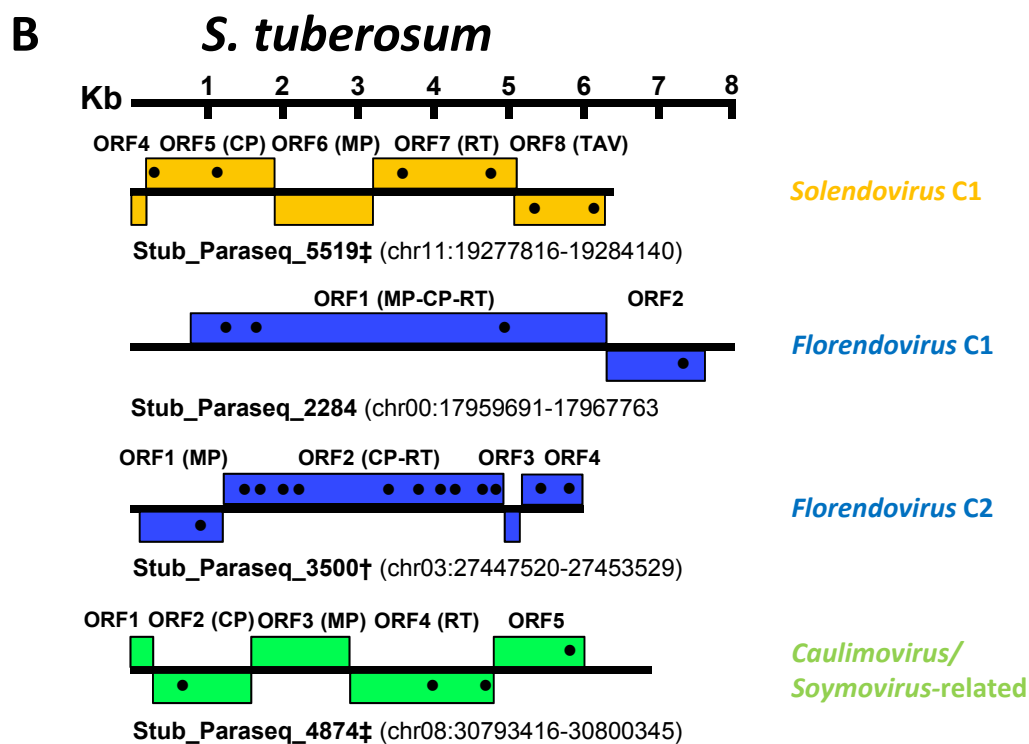
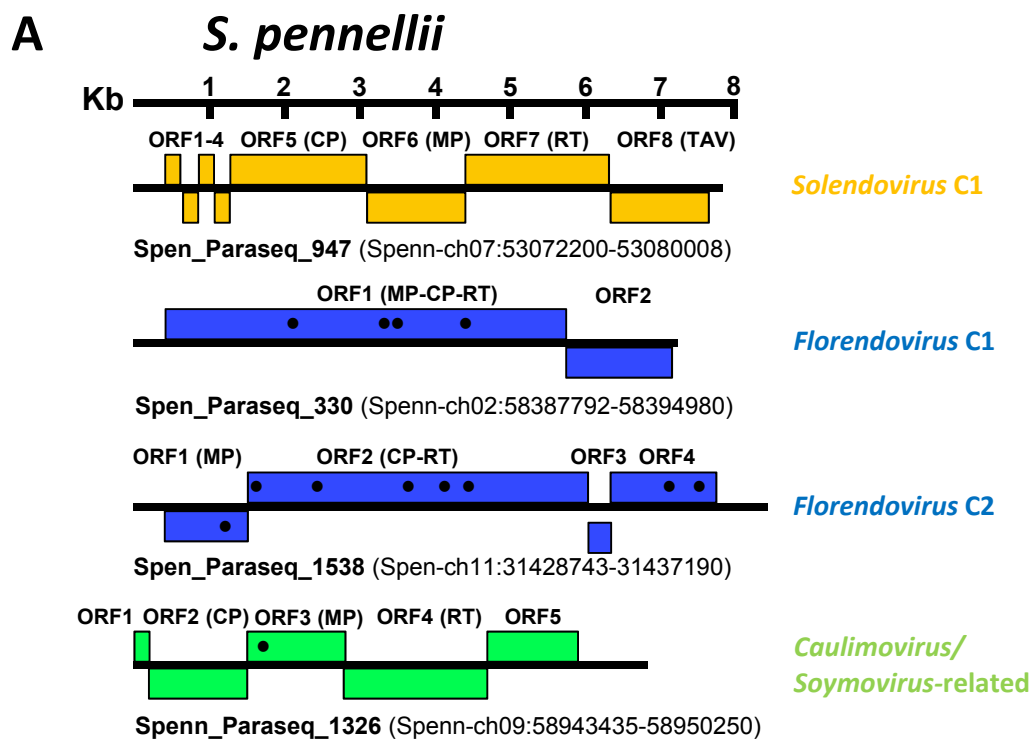
## Supplementary Figure S1

**B**

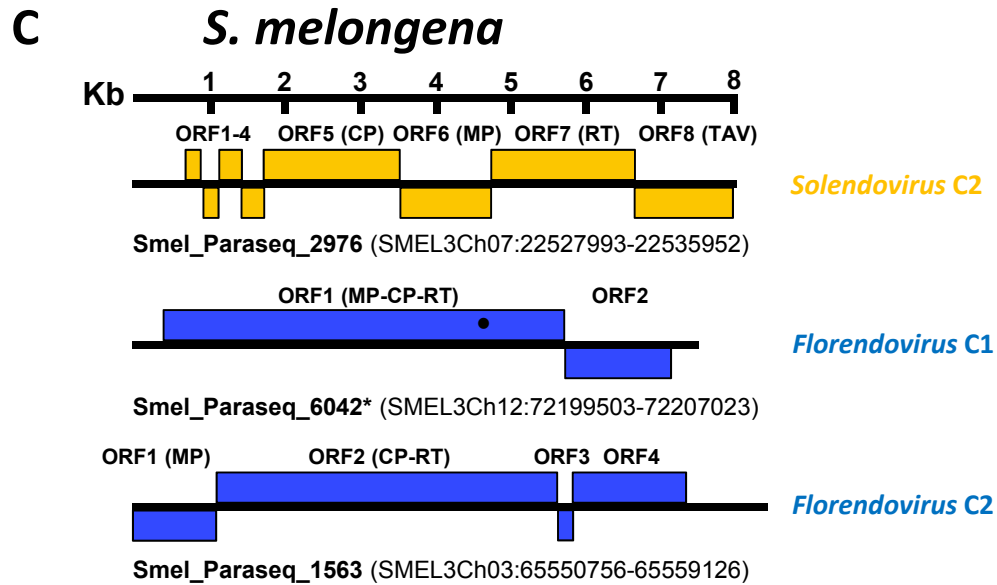


**Supplemental Figure S1:** (A) Global topology of unrooted phylogenetic trees inferred from reverse-transcriptases (RTs) protein sequences with >300 amino-acids (as a conservative threshold to ensure robust confidence in alignment) and their corresponding movement-proteins (MPs), called within recognized pararetroviral sequences in current *S. lycopersicum* (red), *S. pennellii* (blue), *S. tuberosum* (black) and *S. melongena* (violet) genomes. Sixteen different pararetroviruses along with two complete reported endogenized elements from *Florendovirus* (*StubV\_scSt1* and *OsatBV\_compAsc1* from *S. tuberosum* and *Oryza sativa*, respectively) were added (in pink) to expose phylo-group relatedness. Phyletic relationship highlighted for *Solendovirus* (orange; sub-clades C1 and C2), *Florendovirus* (blue; sub-clades C1 and C2) and *Caulimovirus/Soymovirus-related* (green). Others not highlighted (*Badnavirus*, *Tungrovirus*, *Cavemovirus*, *Petuvirus* and *Rosadnavirus*) are presented with black phyletic lines. For simplification, bootstrap support is not shown. f\_TR = denotes sequences bearing recognized flanking tandem-repeats. (B) Global topology of an unrooted phylogenetic tree inferred from reverse-transcriptases (RTs) protein sequences with >300 amino-acids as before, but now with the addition of sequences from representative complete *Solanum* LTR retrotransposons (highlighted cyan) from Copia (*Pseudoviridae*) and Gypsy (*Metaviridae*) superfamilies; other phyletic relationships were highlighted as before. For simplification, here only end labels from pararetroviruses and two complete reported endogenized elements from *Florendovirus* are shown. Note the considered significant ( $\geq 90$ ) bootstrap support separating *Pseudoviridae* and *Metaviridae* from *Caulimoviridae* phyletic lines.

# Supplementary Figure S2



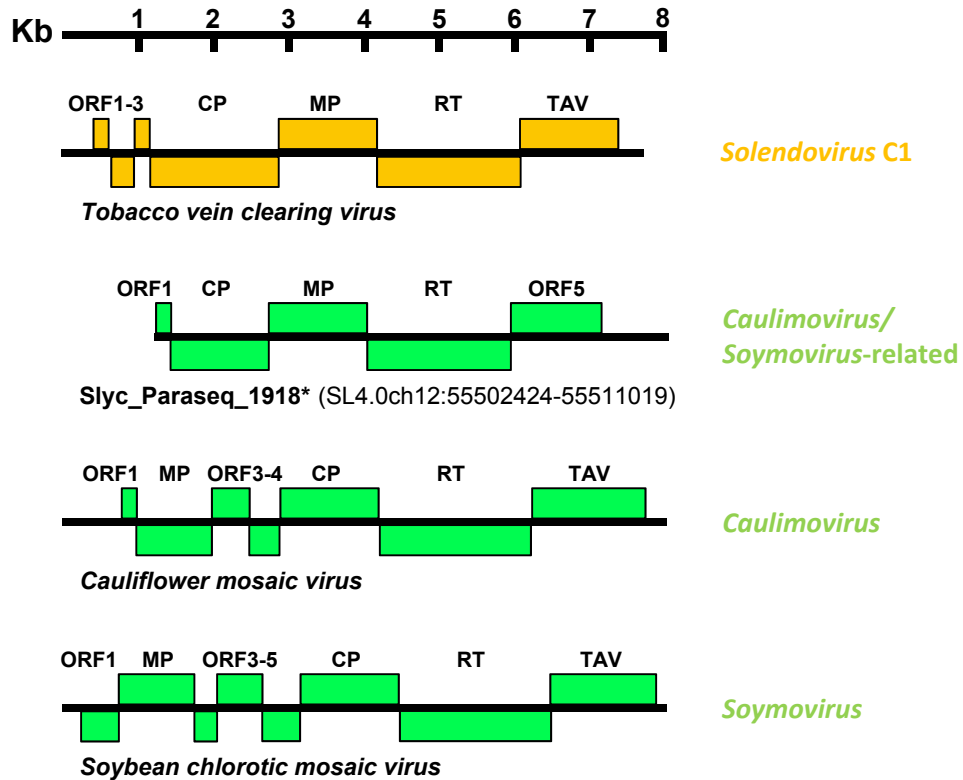
## Supplementary Figure S2



**Supplemental Figure S2:** Schematic representation of exemplary *S. pennellii*, *S. tuberosum* and *S. melongena* EPRVs assembled from distinct phylo-groups. Boxes represent ORFs, with black dots indicating positions with premature stop codons, manually curated using as template other EPRV relatives of the corresponding phylogenetic clade. ORFs encoding conserved domains are indicated between brackets (CP = capsid protein, MP = movement protein, RT = reverse-transcriptase, TAV = transactivator/viroplasm, ORFX = unknown). Coordinates of each EPRV are indicated. Note that elements Stub\_Paraseq\_5519‡, Stub\_Paraseq\_3500†, Stub\_Paraseq\_4874‡ and Smel\_Paraseq\_6042\* were not listed among the non-truncated EPRVs (Supplemental Table S2); but were assembled and added here to reflect the presence of their corresponding phylo-groups. Stub\_Paraseq\_5519‡ and Stub\_Paraseq\_4874‡ did not achieve the stringent thresholds of identity/alignment set for elements in Supplemental Table S2. Stub\_Paraseq\_3500† was undersized ruled out from Supplemental Table S2, because the automatically-generated original lacked an unrecognized immediately downstream 1008 bp fragment comprising ORF3 y la ORF4. On the other hand, Smel\_Paraseq\_6042\* is a portion of an original parasequence which was ruled out due to oversize, but involved rearranged sequences additionally extending the otherwise complete non-truncated element here presented. Note that we failed to document the occurrence of the novel *Caulimovirus*/*Soymovirus*-related family in *S. melongena*.

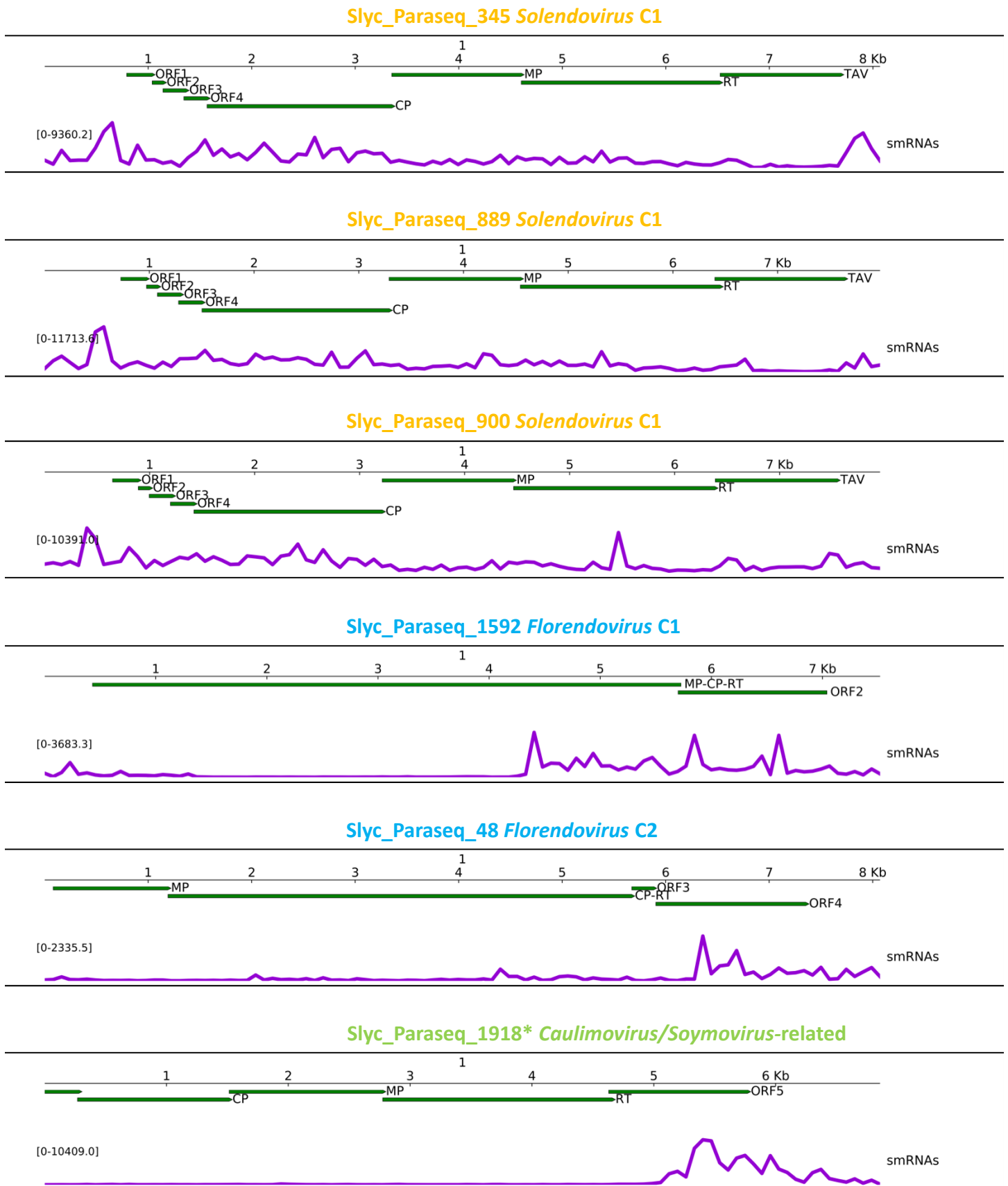


## Supplementary Figure S3



**Supplemental Figure S3:** Schematic comparative coding organization between *Solendovirus*, *Caulimovirus*, *Soymovirus*, and the novel *Caulimovirus/Soymovirus*-related family. ORFs encoding conserved domains are indicated between brackets (CP = capsid protein, MP = movement protein, RT = reverse-transcriptase, TAV = transactivator/viroplasmin, ORFX = unknown). Note that the core CP/MP/RT in novel *Caulimovirus/Soymovirus*-related elements is comparable to that of *Solendovirus* but differs from the MP/CP/RT core of *Caulimovirus* and *Soymovirus* (in spite of RT phylogenetic relatedness), supporting the idea that they belong to a not-yet-described genus. S lyc\_Paraseq\_1918\* is a portion of an original parasequence which presented a rearrangement at the 5' end, additionally extending the otherwise complete non-truncated element here presented.

## Supplementary Figure S4



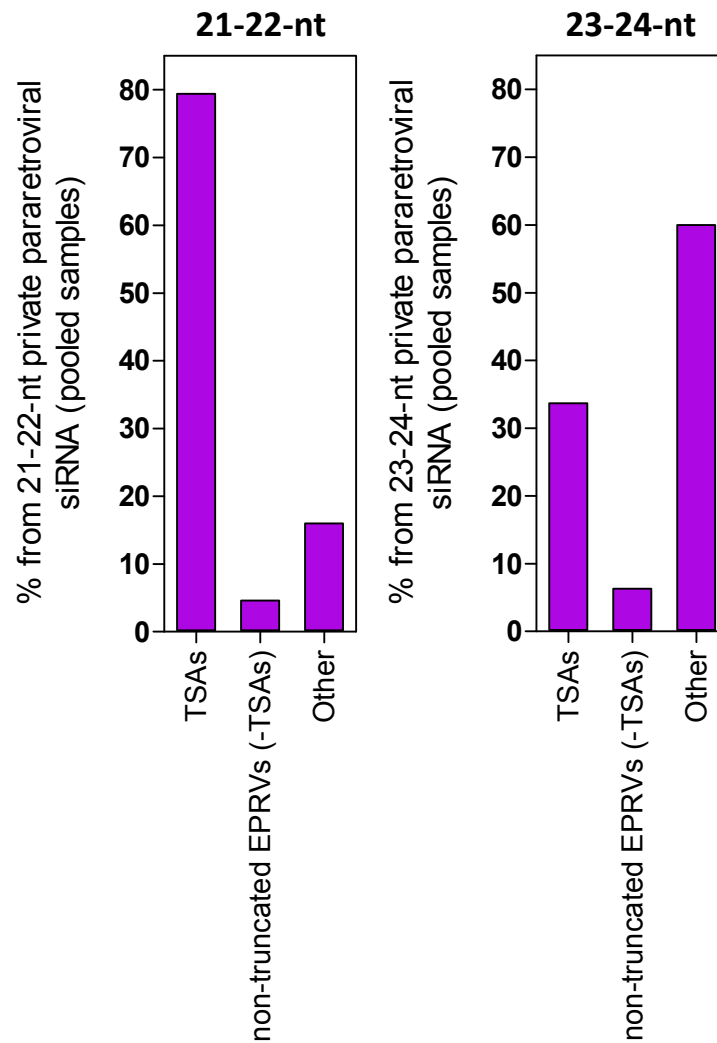
**Supplemental Figure S4:** *S. lycopersicum* exemplary assembled EPRVs as templates for mapping endogenous pararetroviral siRNAs, collapsed from different tissue libraries. Mayor phylo-groups represented are *Solendovirus* (sub-clade C1), *Florendovirus* (sub-clades C1 and C2) and novel *Caulimovirus/Soymovirus*-related. Inferred coding sequences are shown in green on top (CP = capsid protein, MP = movement protein, RT = reverse-transcriptase, TAV = transactivator/viroplasin, ORFX = unknown).

## Supplementary Figure S5

Tissue	smRNA library size (private 18-25nt)
Shoot_2wk	17529897
Leaf_4wk_young	13481240
Leaf_4wk_green	12758118
Leaf_4wk_yellow	10049315
Pollen_meiotic_tetrad	11206501
Pollen_postmeiotic_microspores	7031550
Pollen_mature_bicellular	4037015
Fruit_ripe	26454973

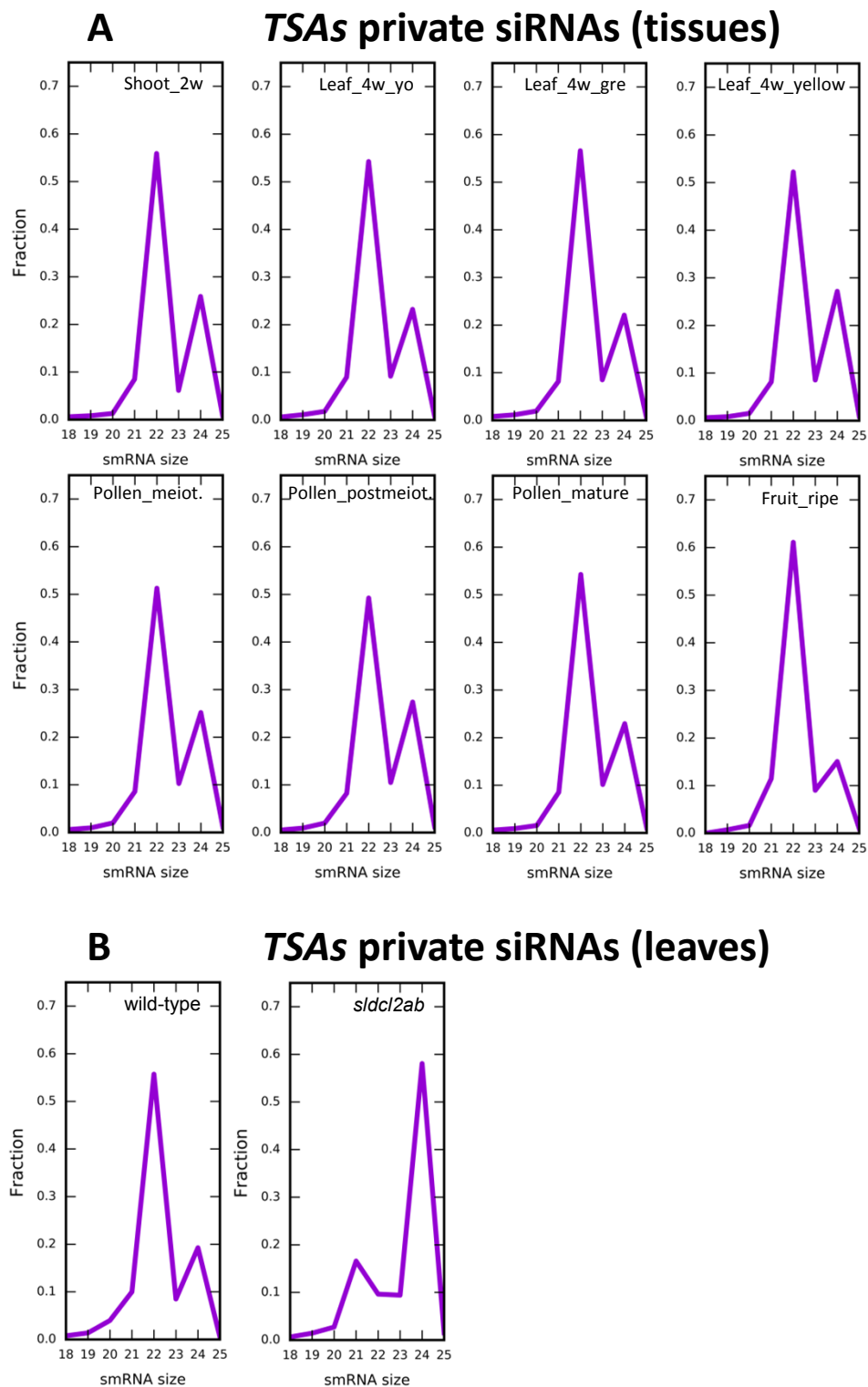
**Supplemental Figure S5:** Private 18-25-nt siRNA library sizes from eight *S. lycopersicum* examined tissues.

## Supplementary Figure S6



**Supplemental Figure S6:** Percentages of certain private siRNA sizes from pooled tissue samples mapping to annotated *S. lycopersicum* pararetroviral-related sequences. *TSAs* = transcriptionally-competent siRNA areas; non-truncated EPRVs (-*TSAs*) = recognized non-truncated EPRVs not belonging to any *TSA*; Other = remaining EPRV-related sequences not included in the previous groups, comprising historical remnants.

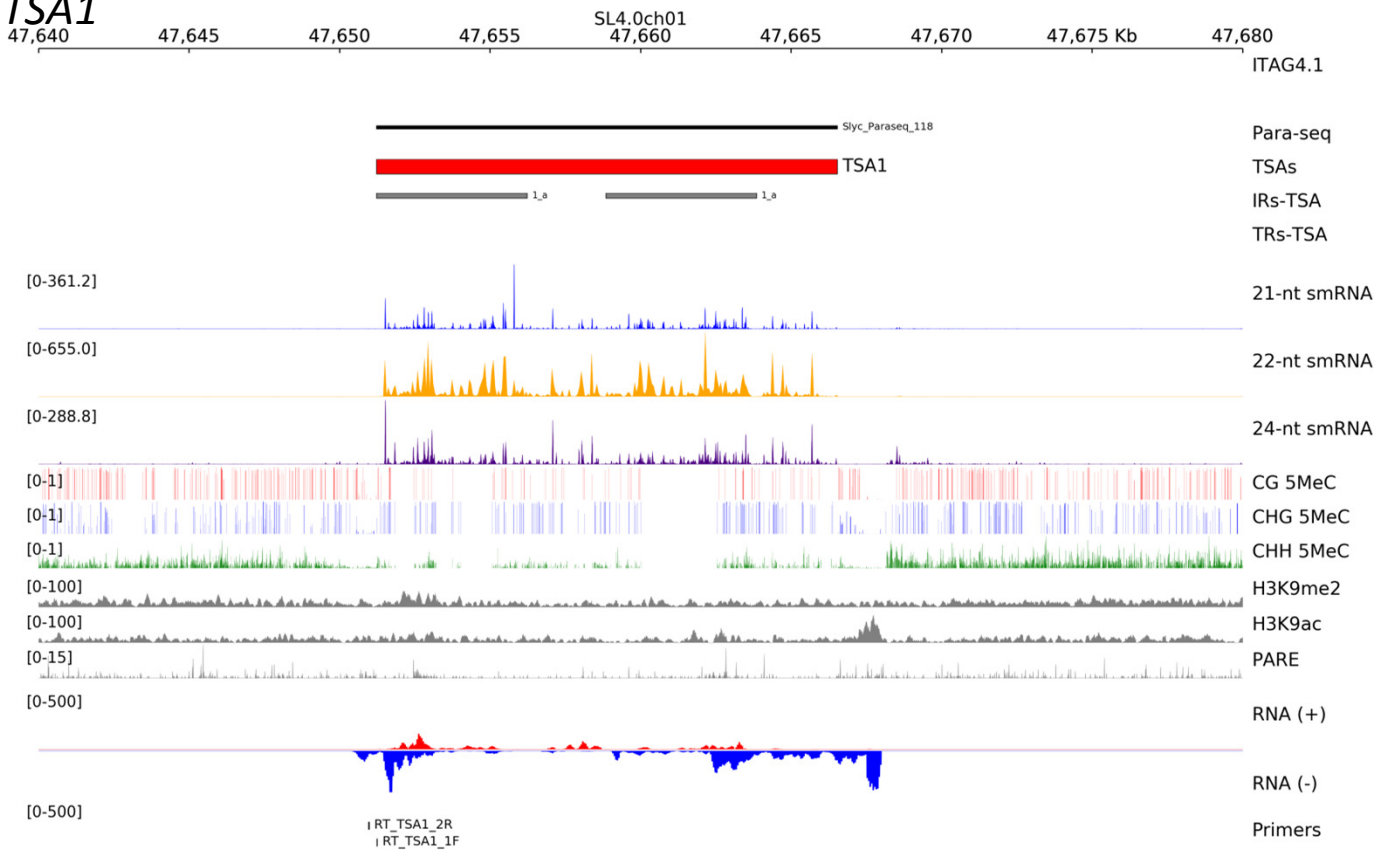
## Supplementary Figure S7



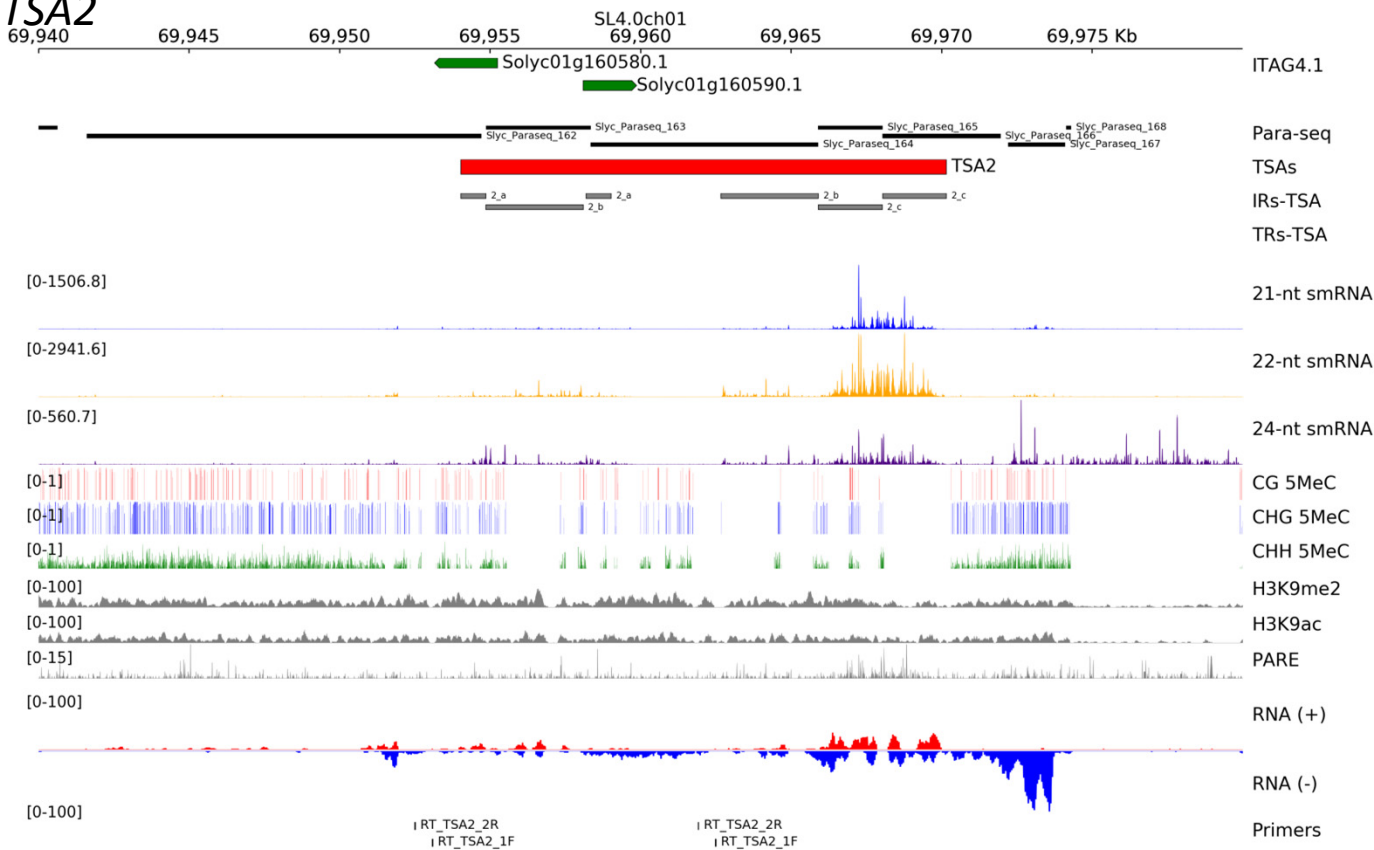
**Supplemental Figure S7:** Private siRNA size profiles mapping to *S. lycopersicum* TSAs: (A) Across eight examined tissues; and (B) in leaves of wild-type (left) and *slacl2ab* mutant (right).

# Supplementary Figure S8

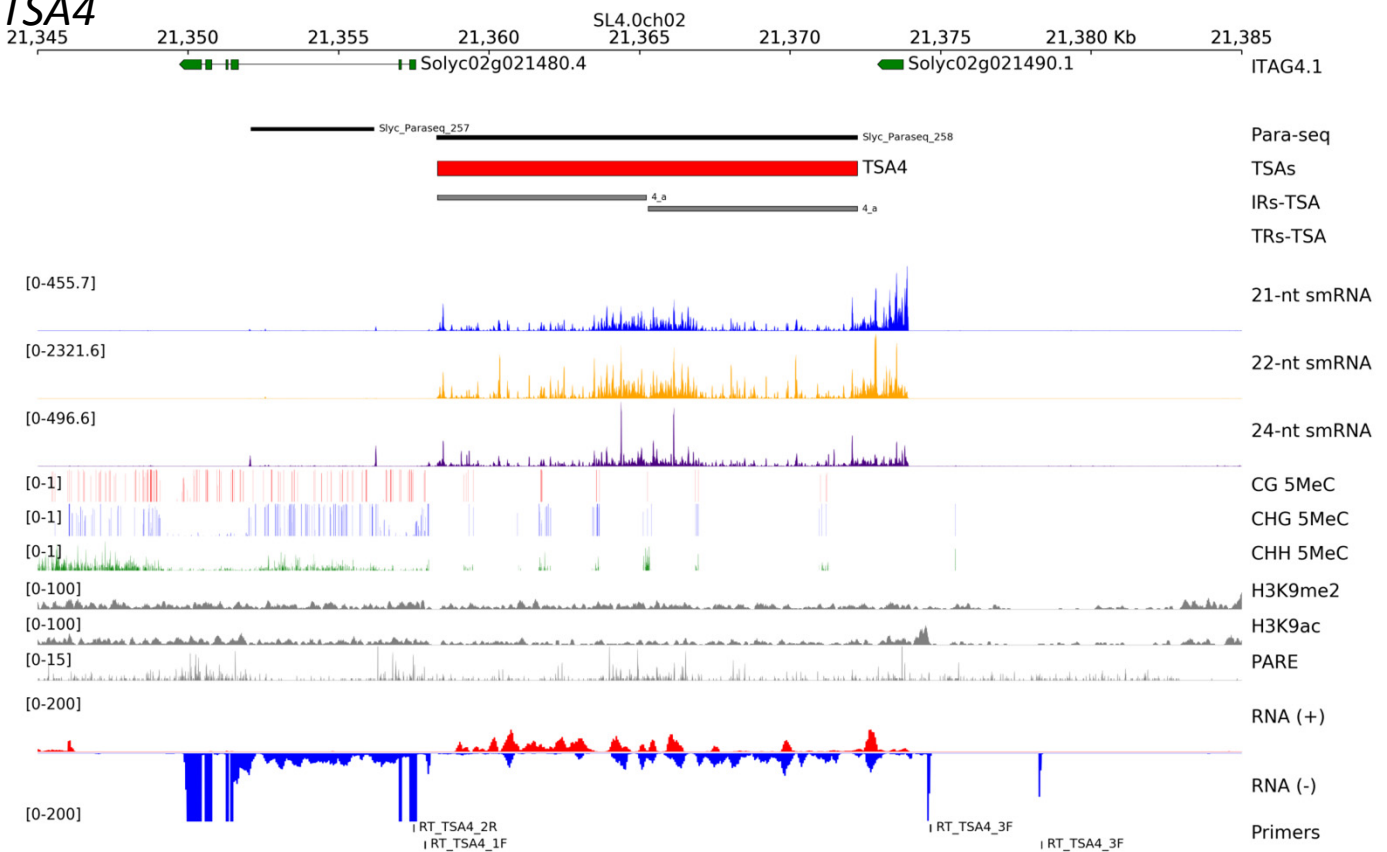
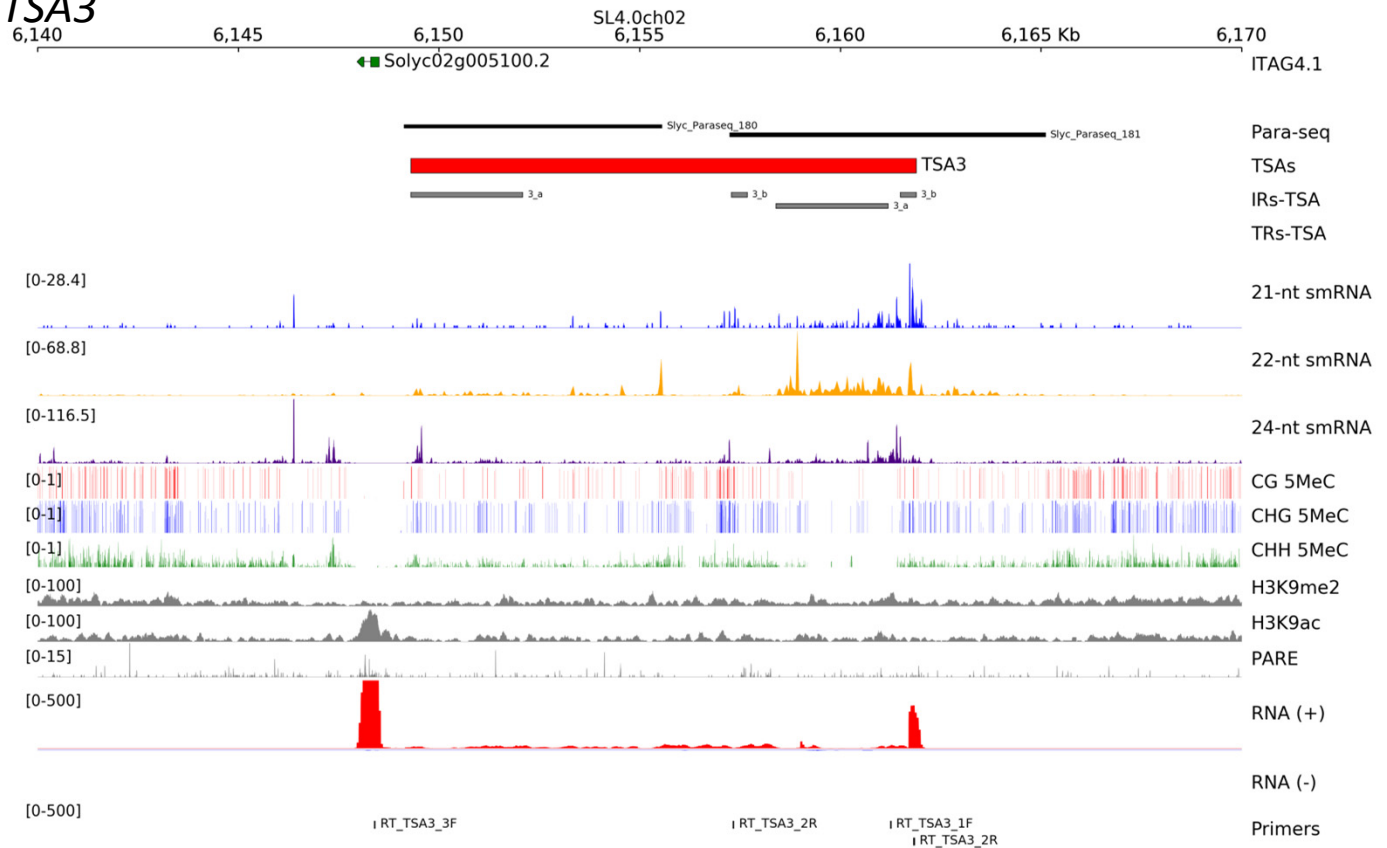
## TSA1



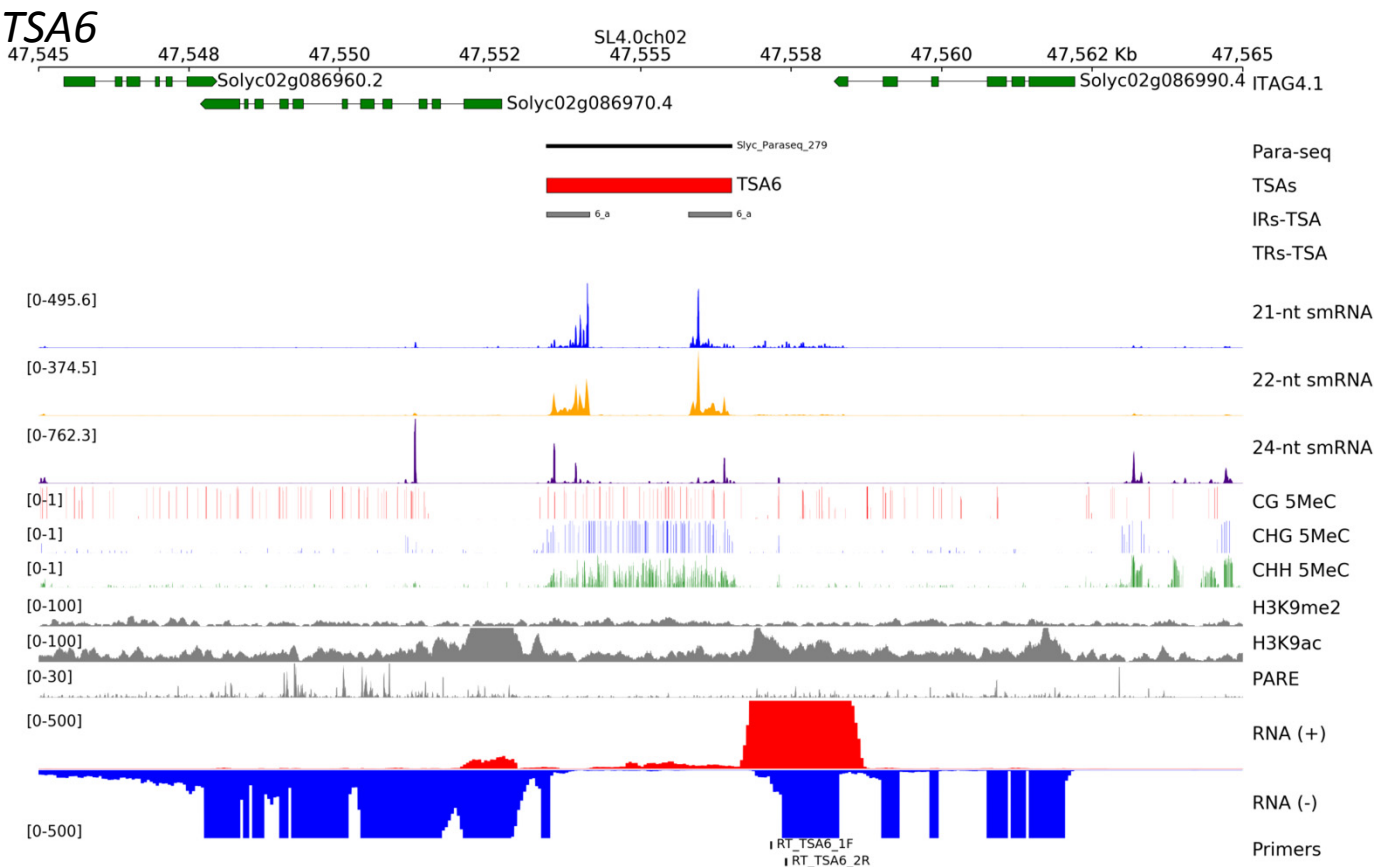
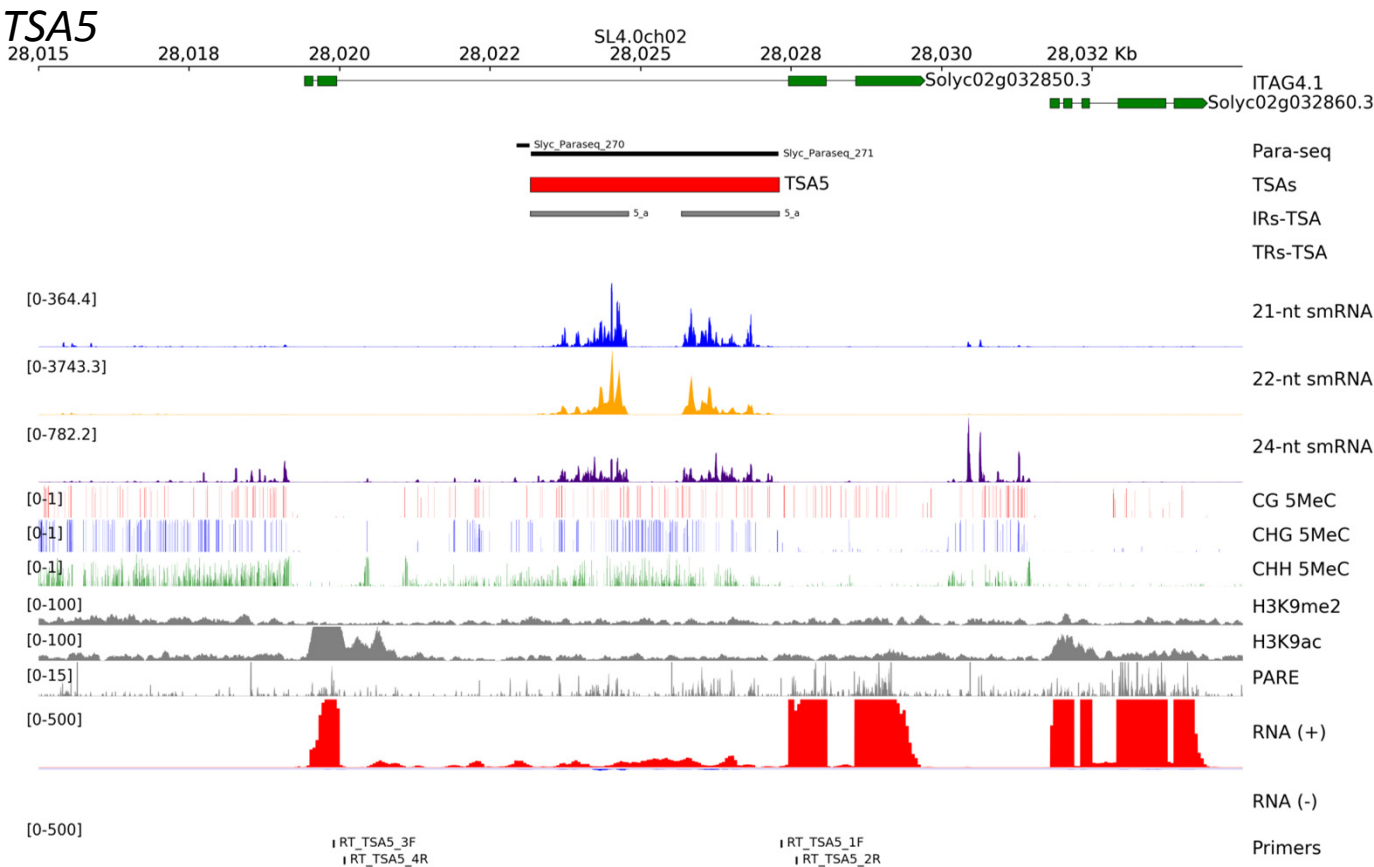
## TSA2



## Supplementary Figure S8



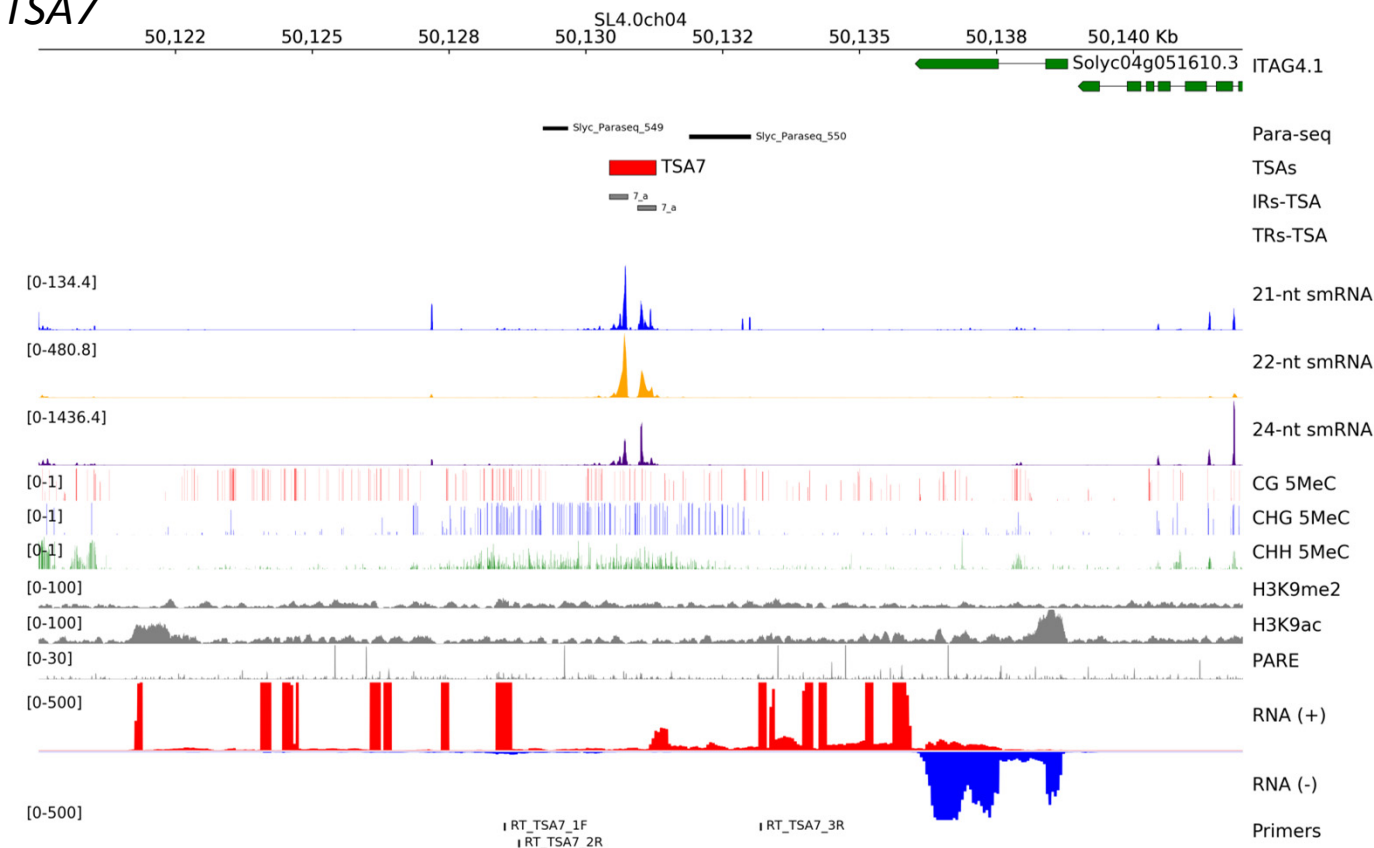
Supplementary Figure S8



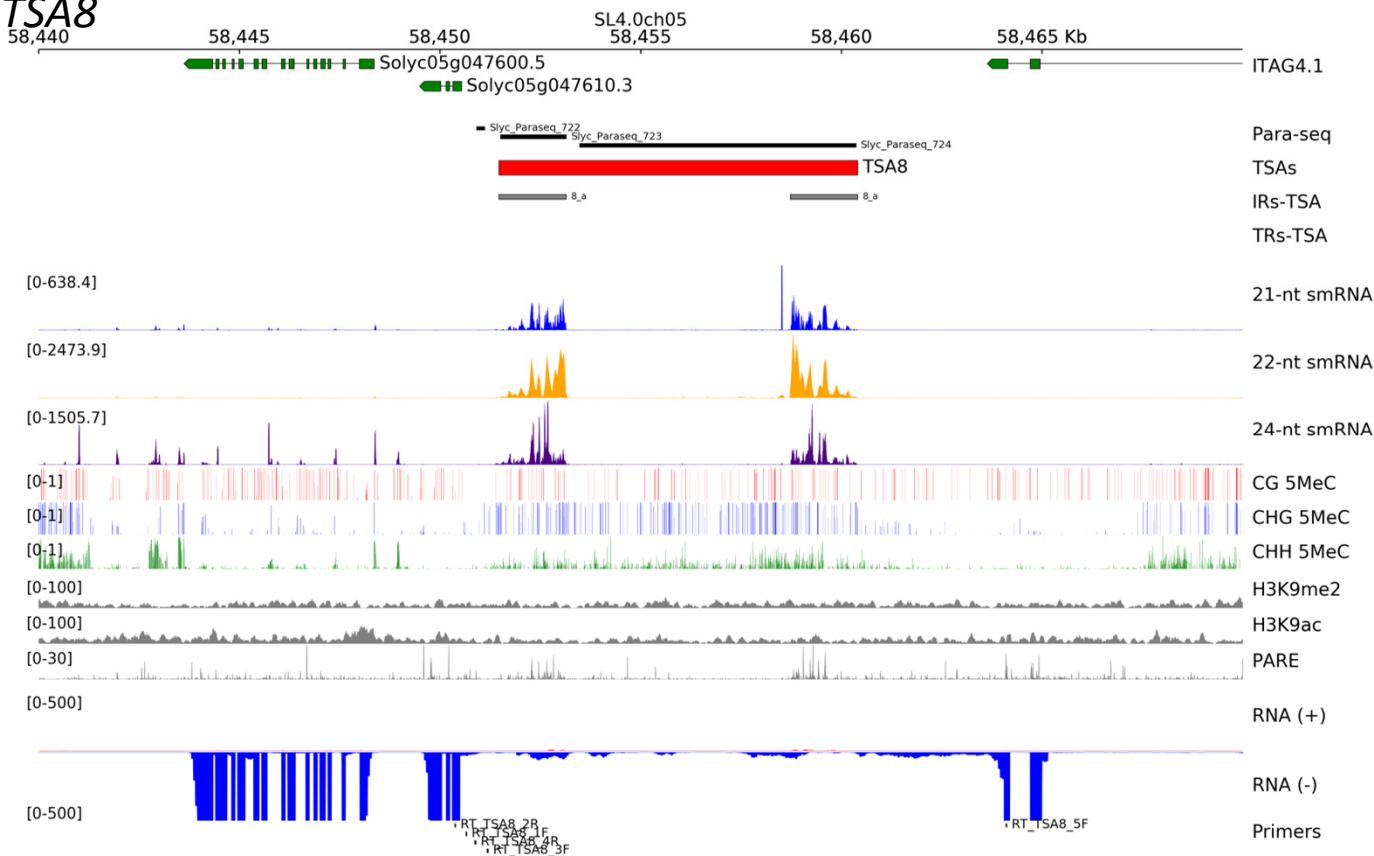


Supplementary Figure S8

TSA7

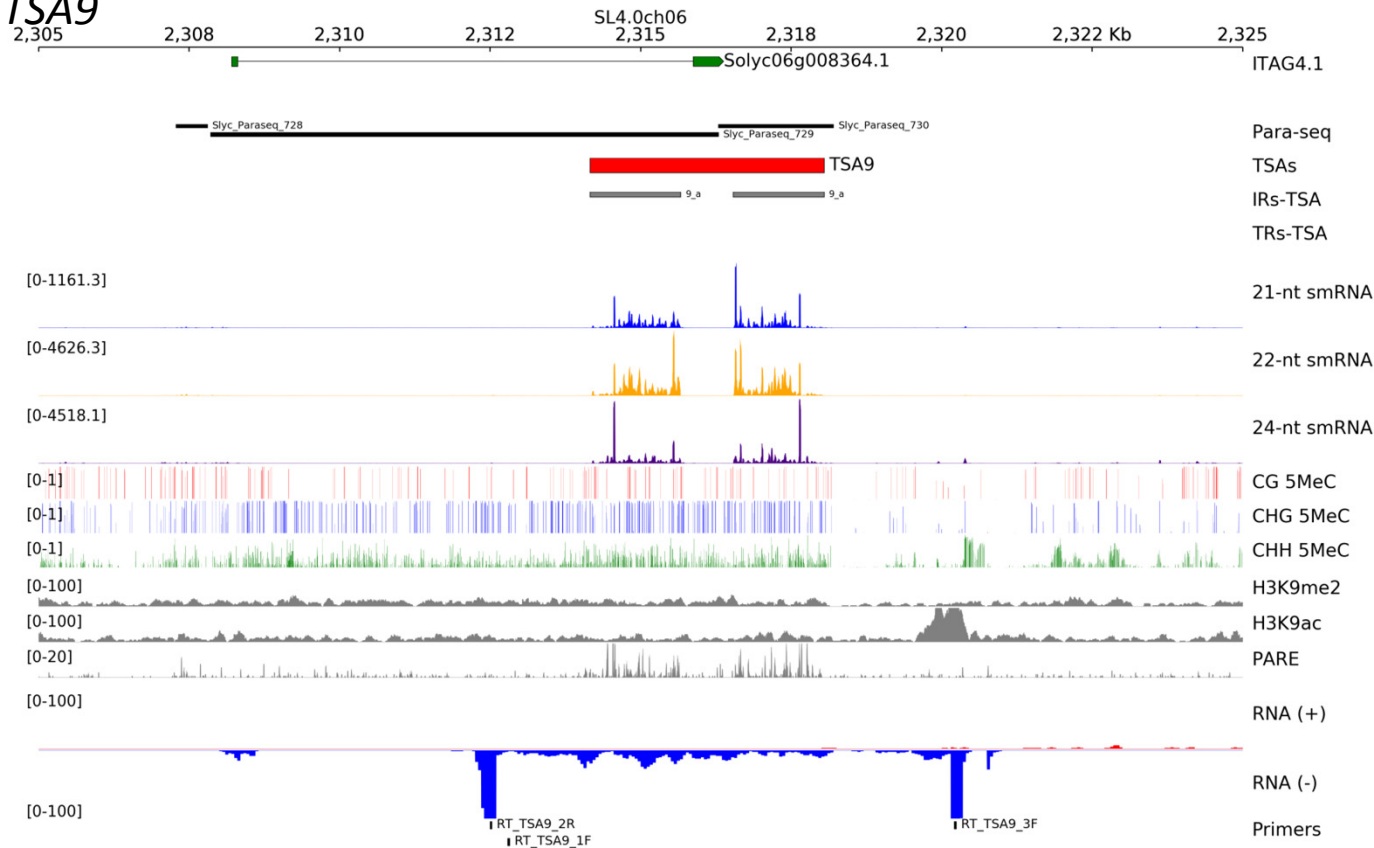


TSA8

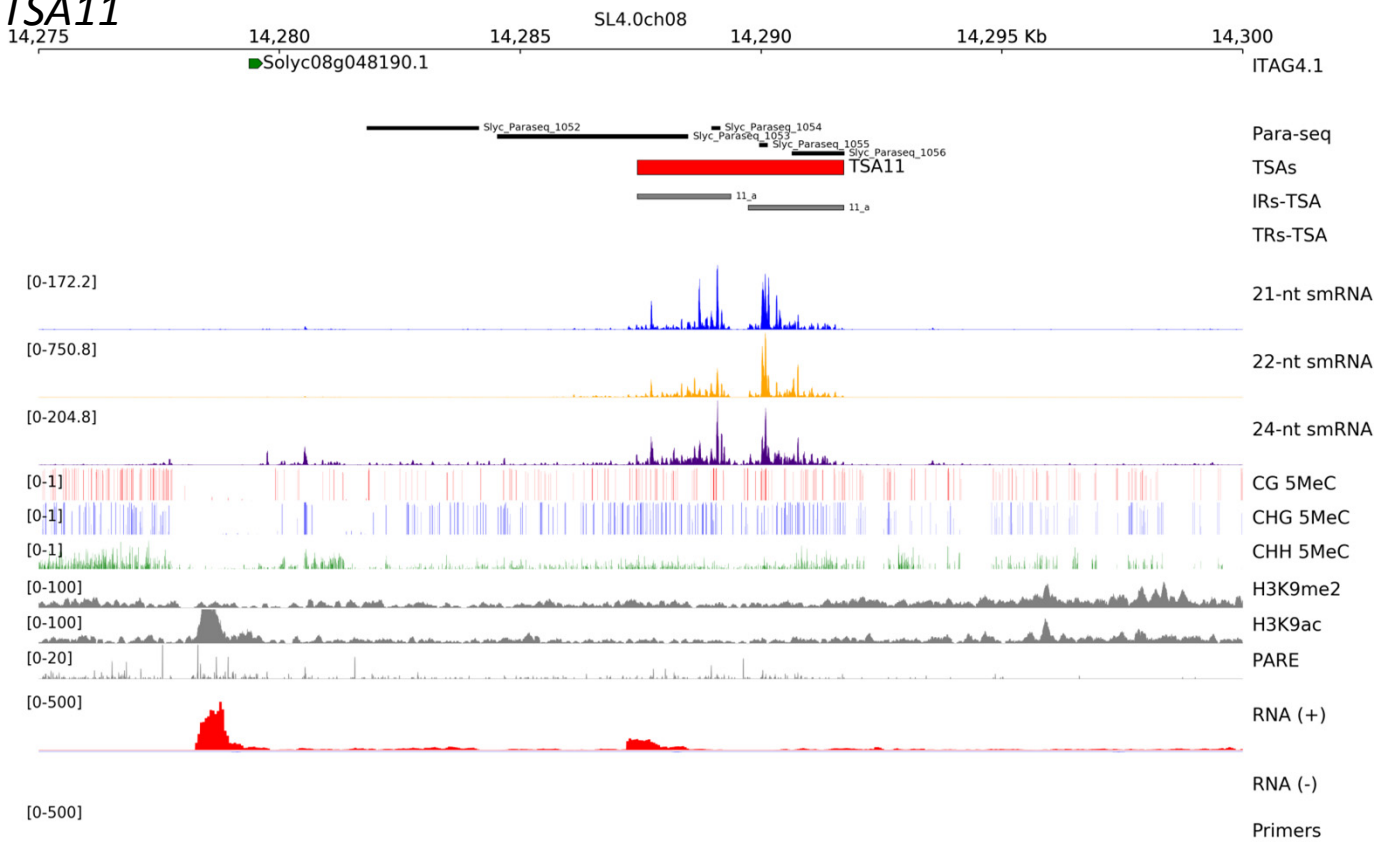


Supplementary Figure S8

TSA9

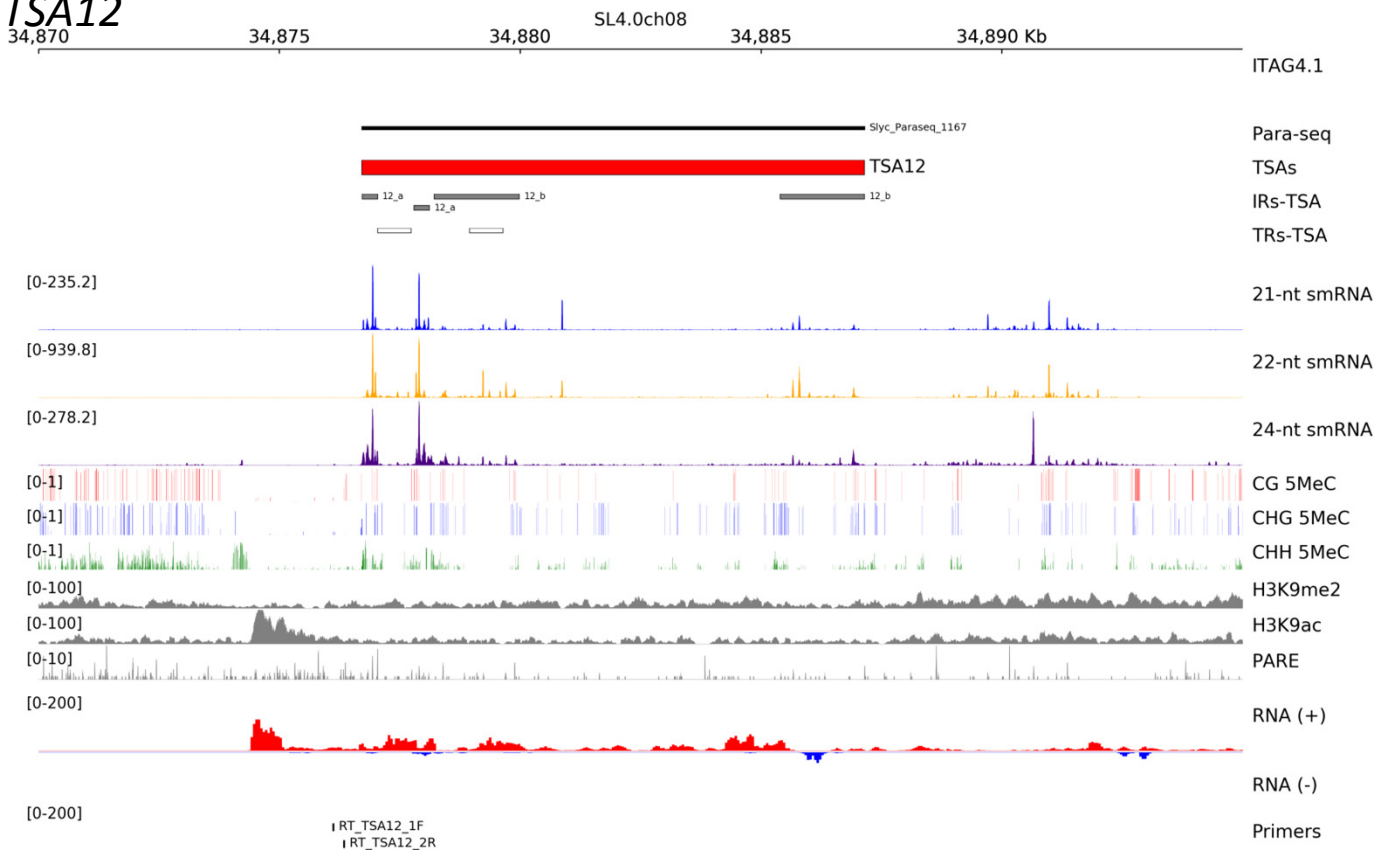


TSA11

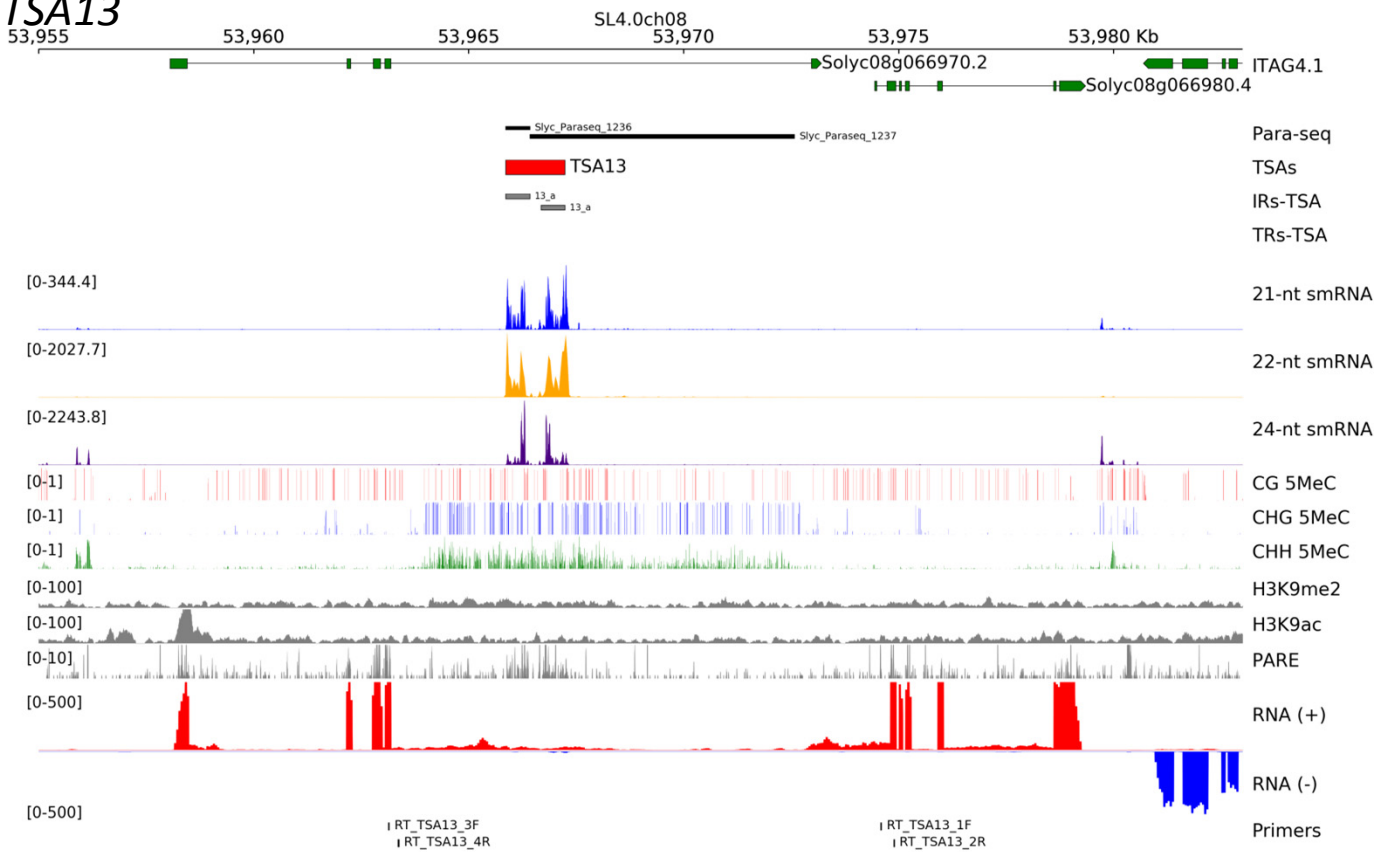


Supplementary Figure S8

TSA12

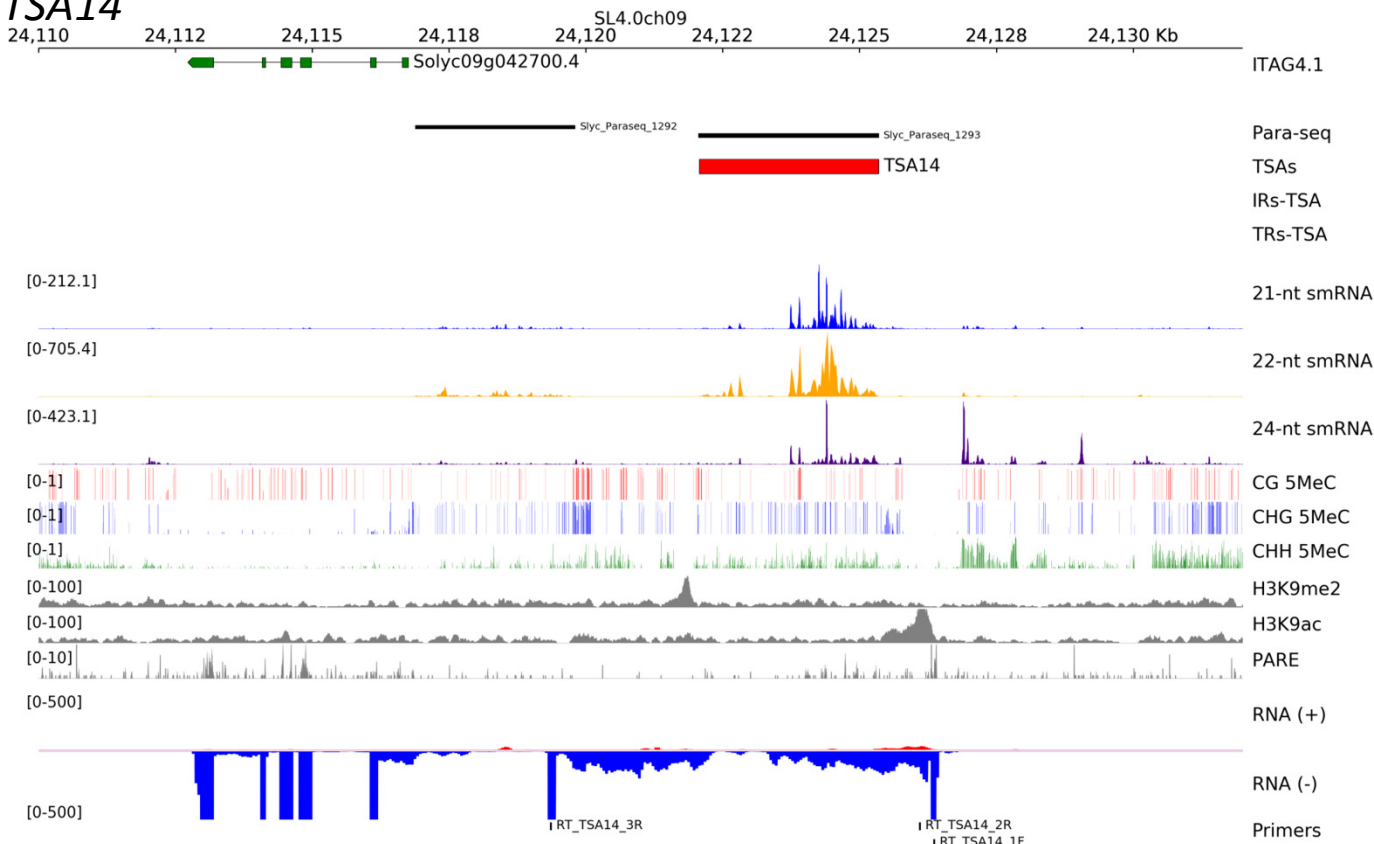


TSA13

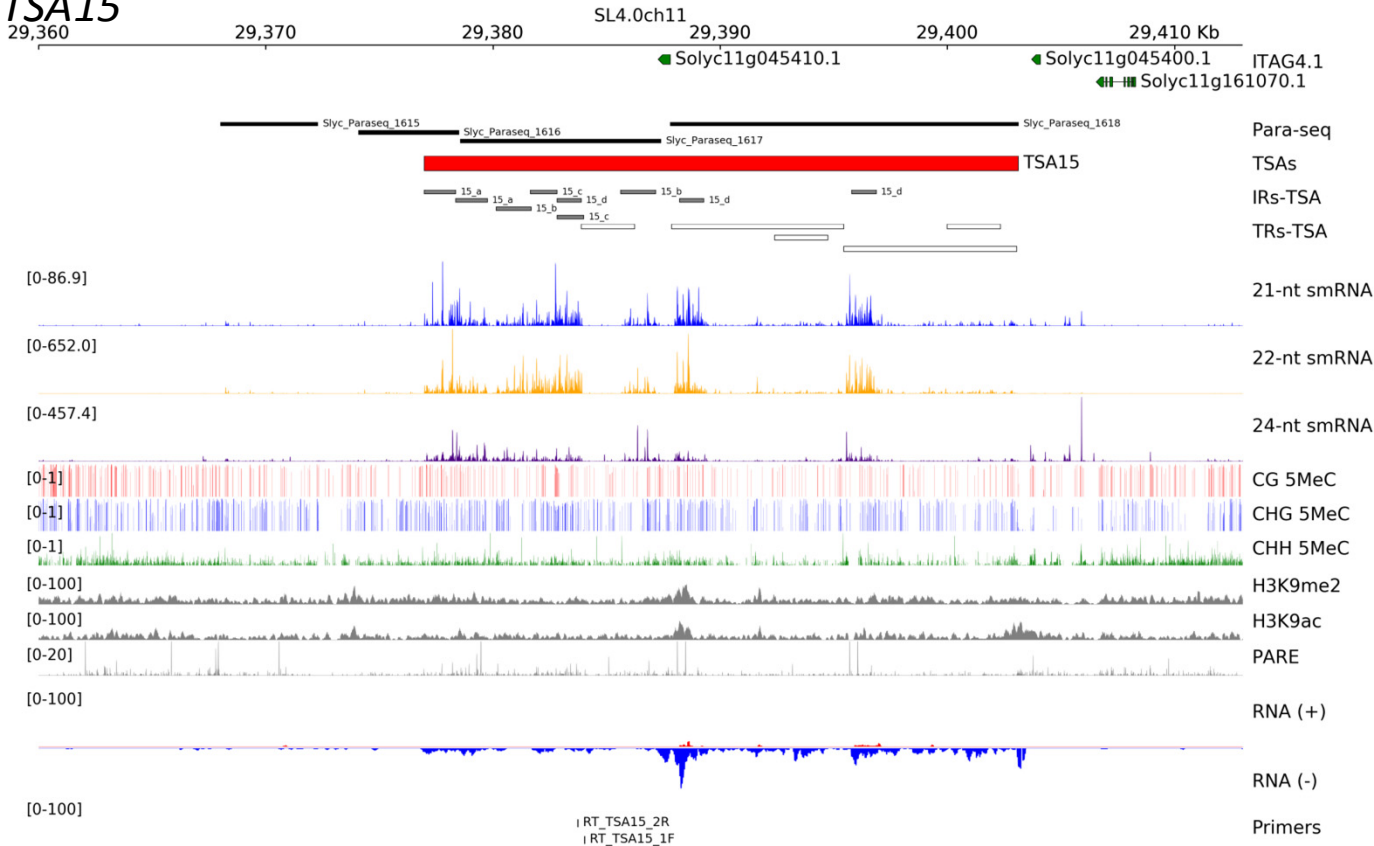


Supplementary Figure S8

TSA14

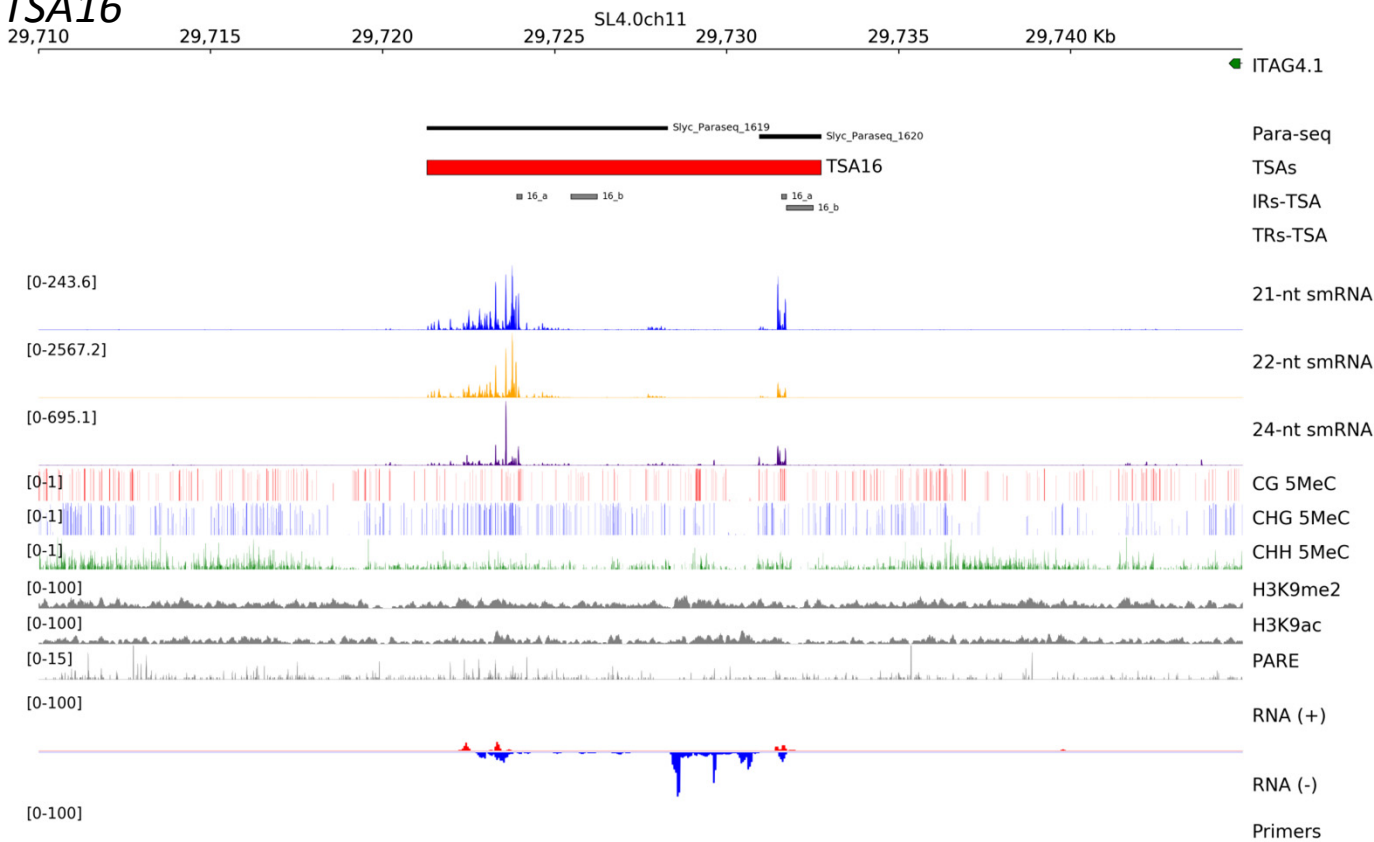


TSA15

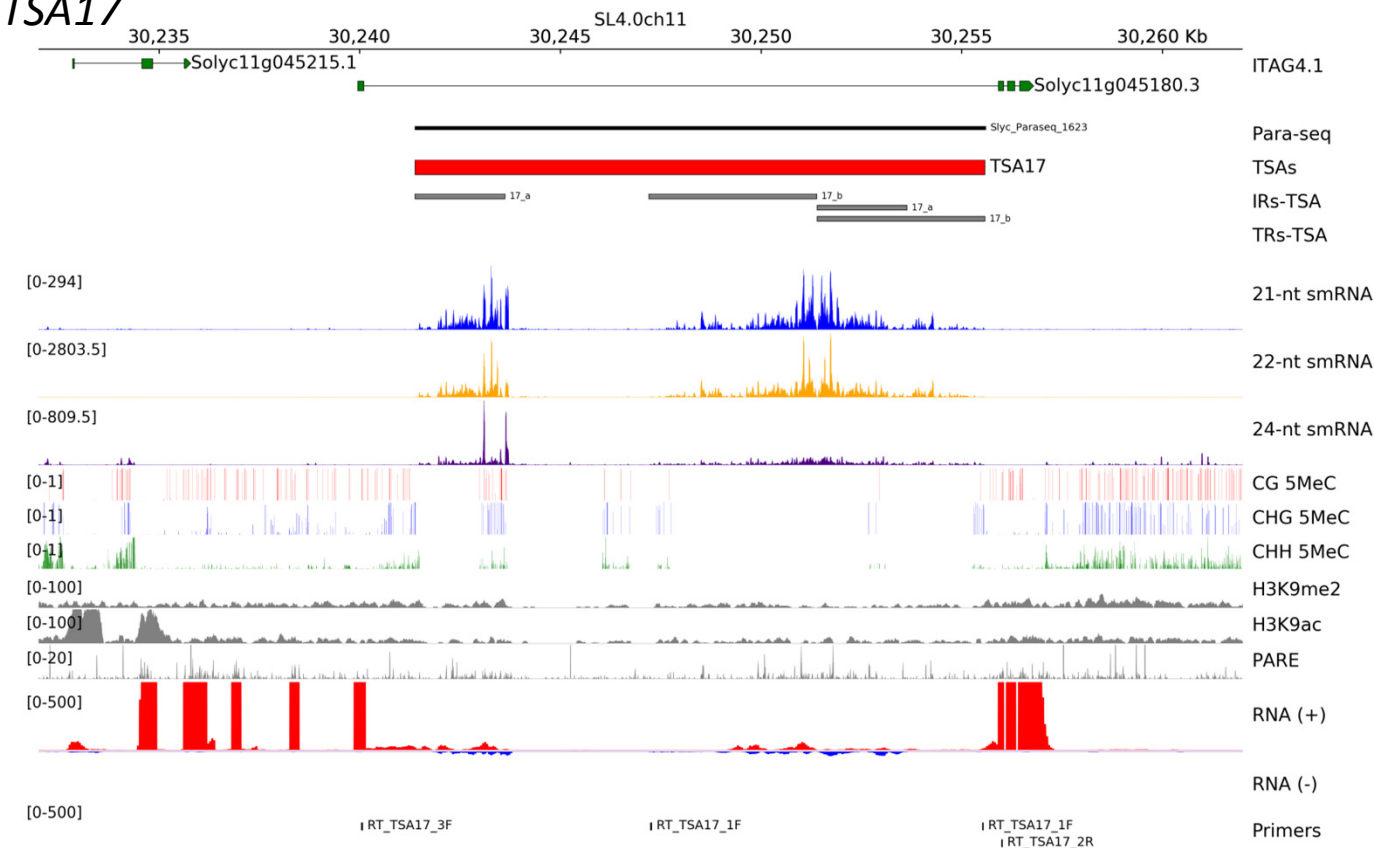


Supplementary Figure S8

TSA16



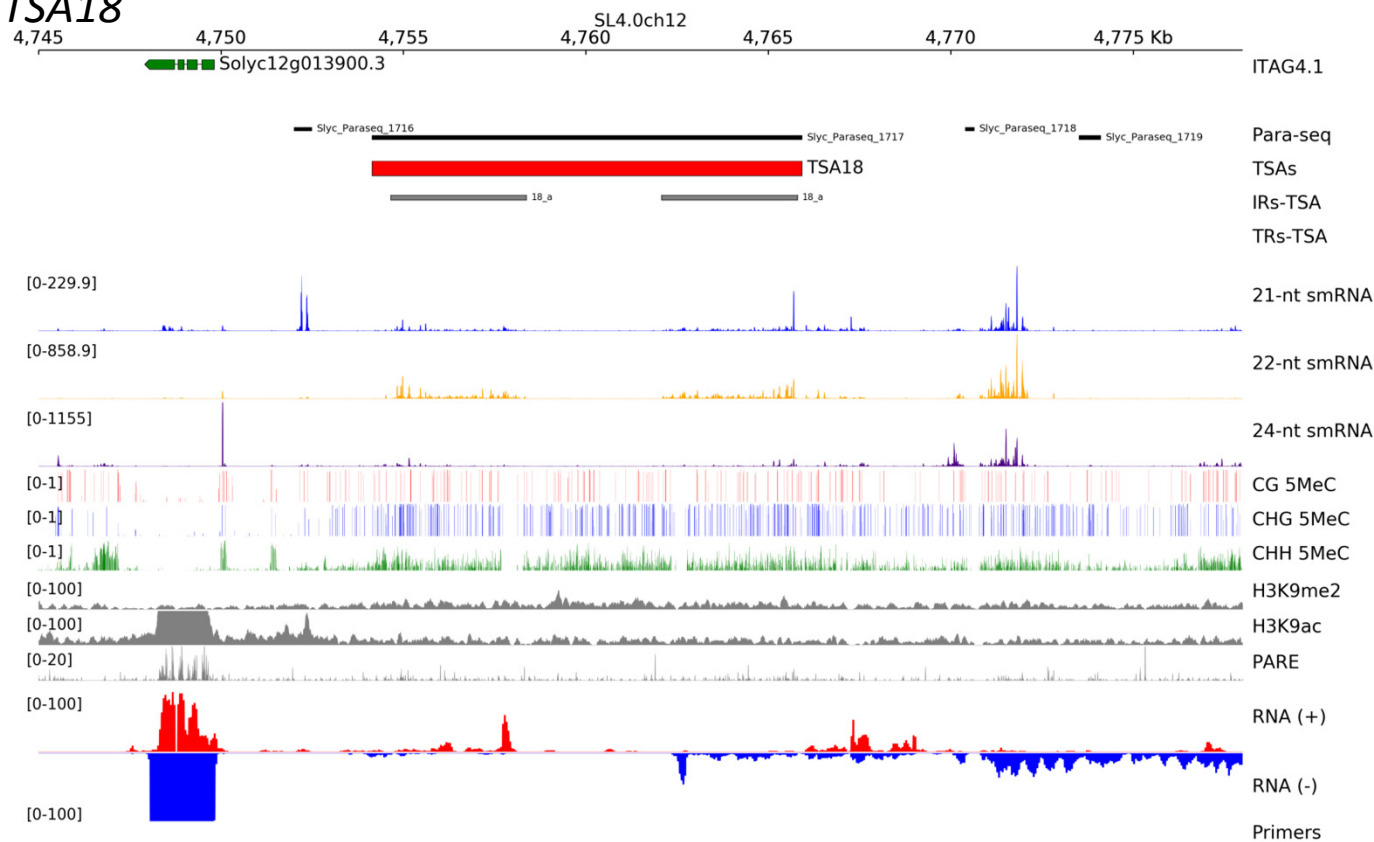
TSA17



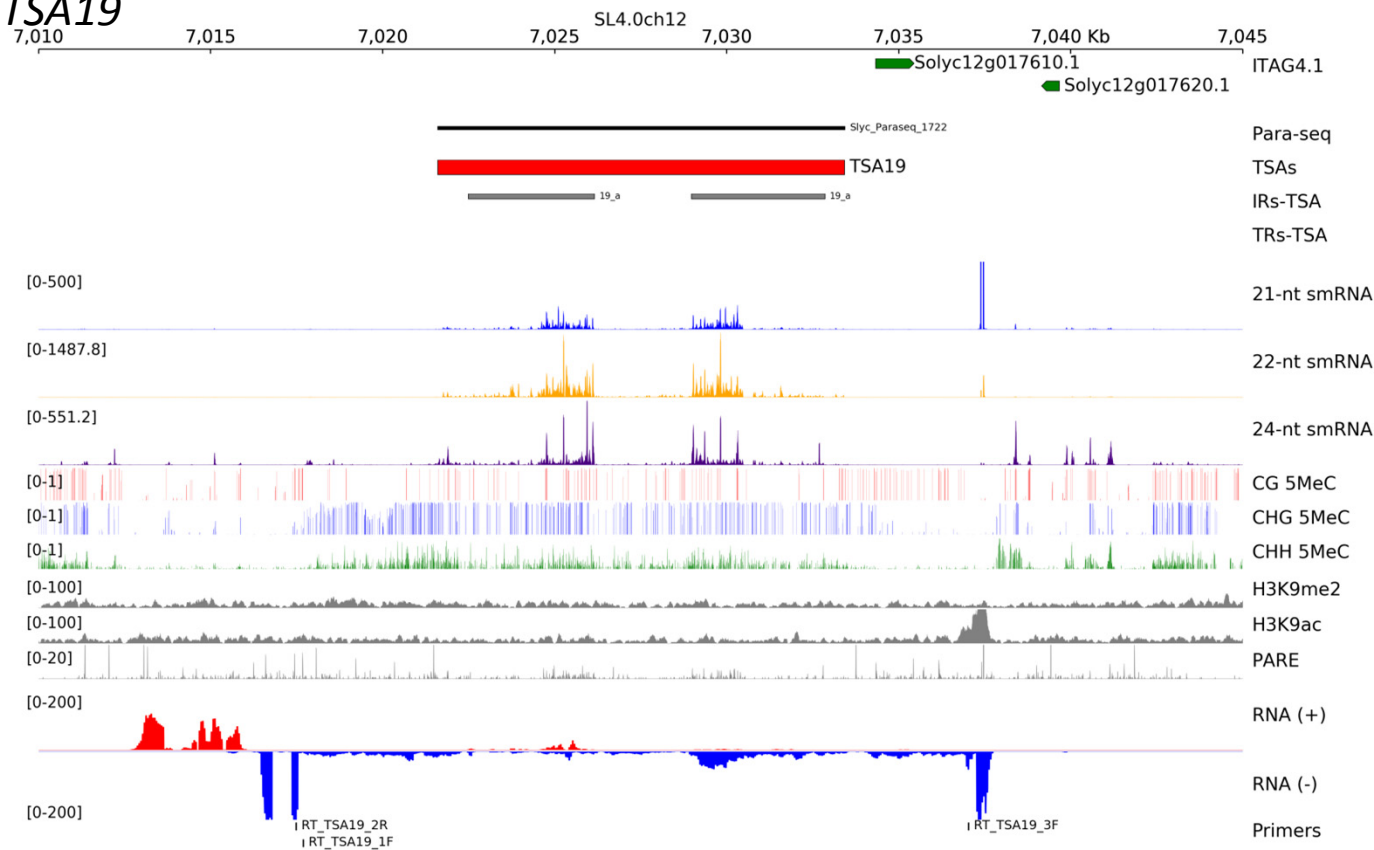


Supplementary Figure S8

TSA18

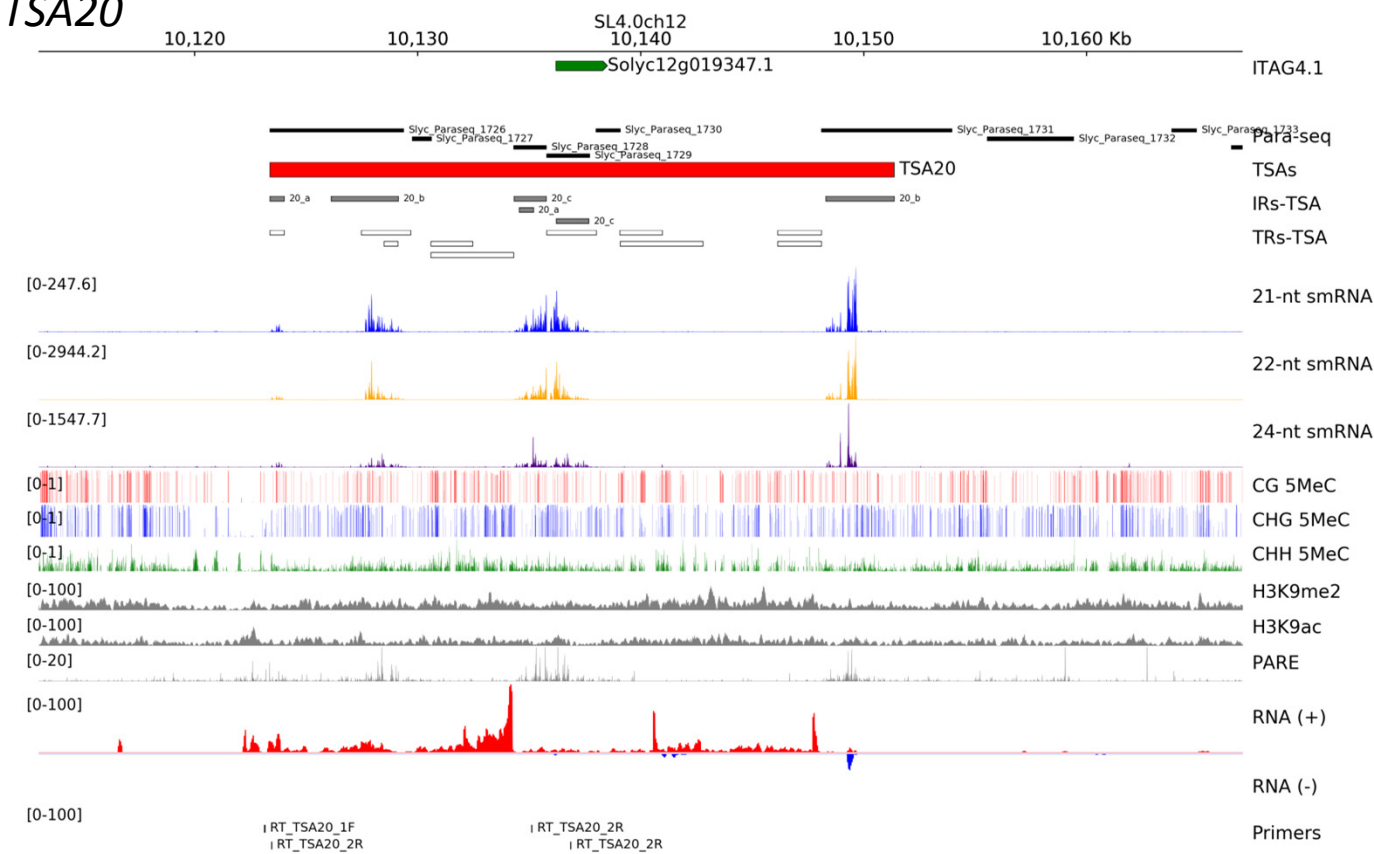


TSA19

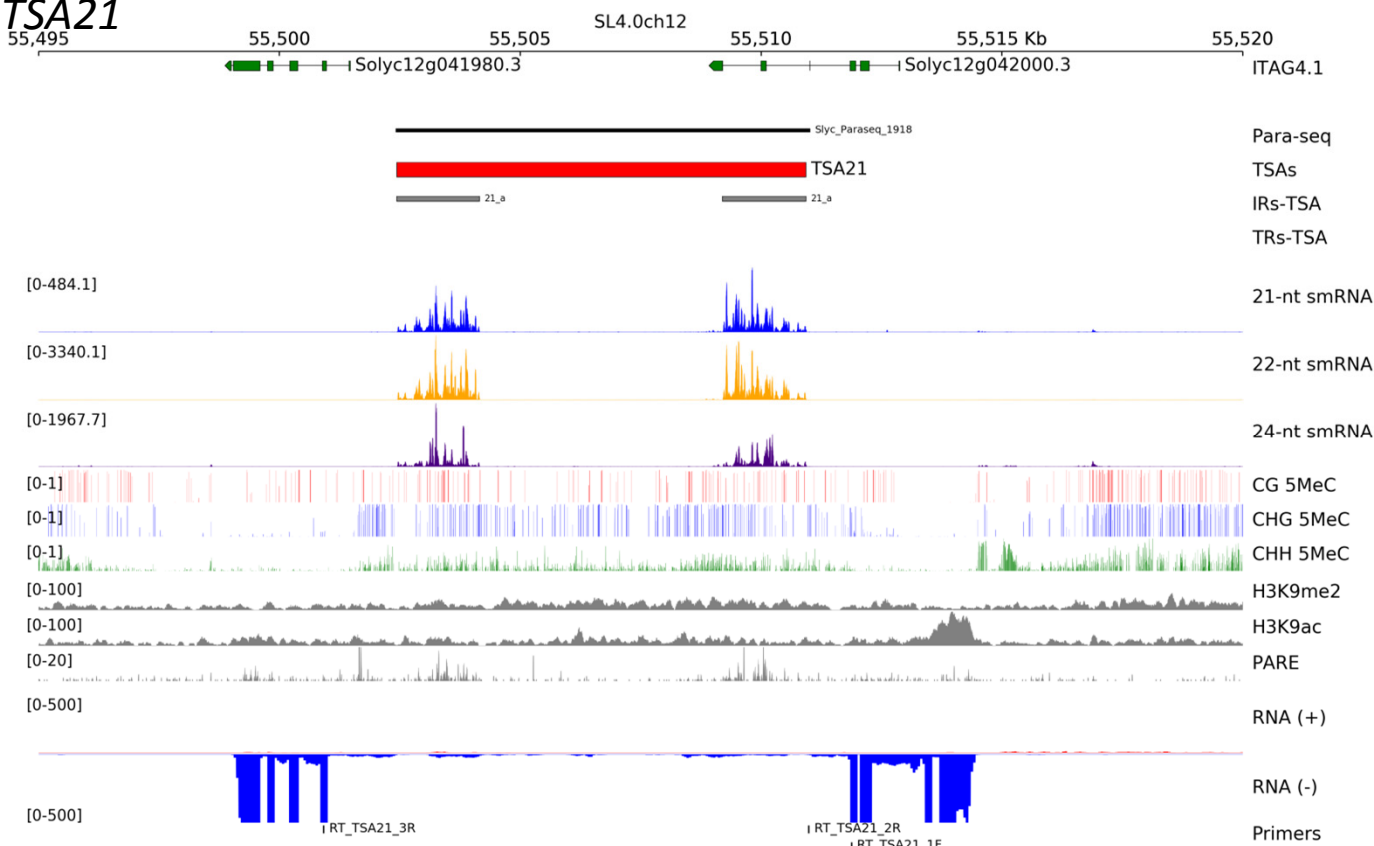


Supplementary Figure S8

TSA20

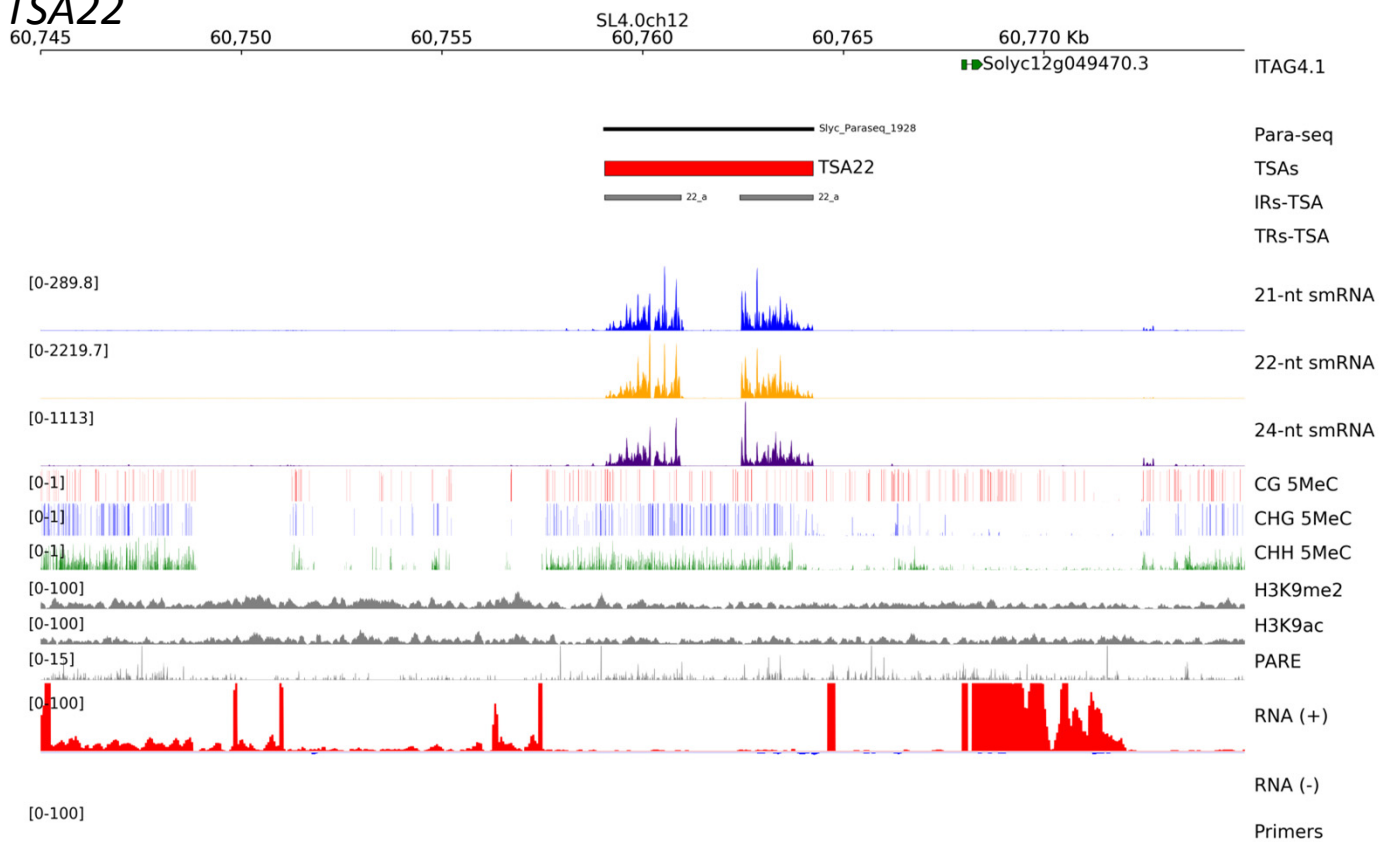


TSA21



# Supplementary Figure S8

**TSA22**



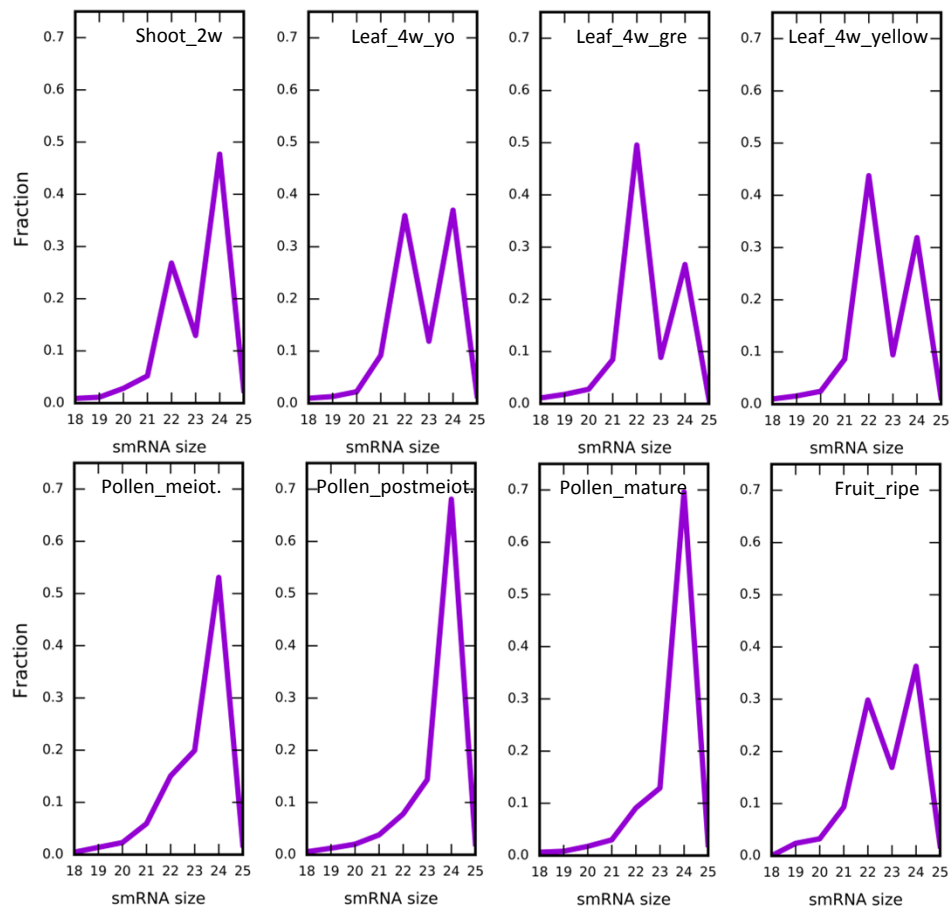
**Supplemental Figure S8:** Genome-browsers around *S. lycopersicum* TSA coordinates. Data visualized as in Figure 3A. Note that RNA-seq signals in several instances suggest coding genetic features incorrectly annotated in current ITAG4.1. Hence, *TSA4*, *TSA7*, *TSA8*, *TSA14*, *TSA21* and *TSA22* were assumed to occur intronically.



## Supplementary Figure S9

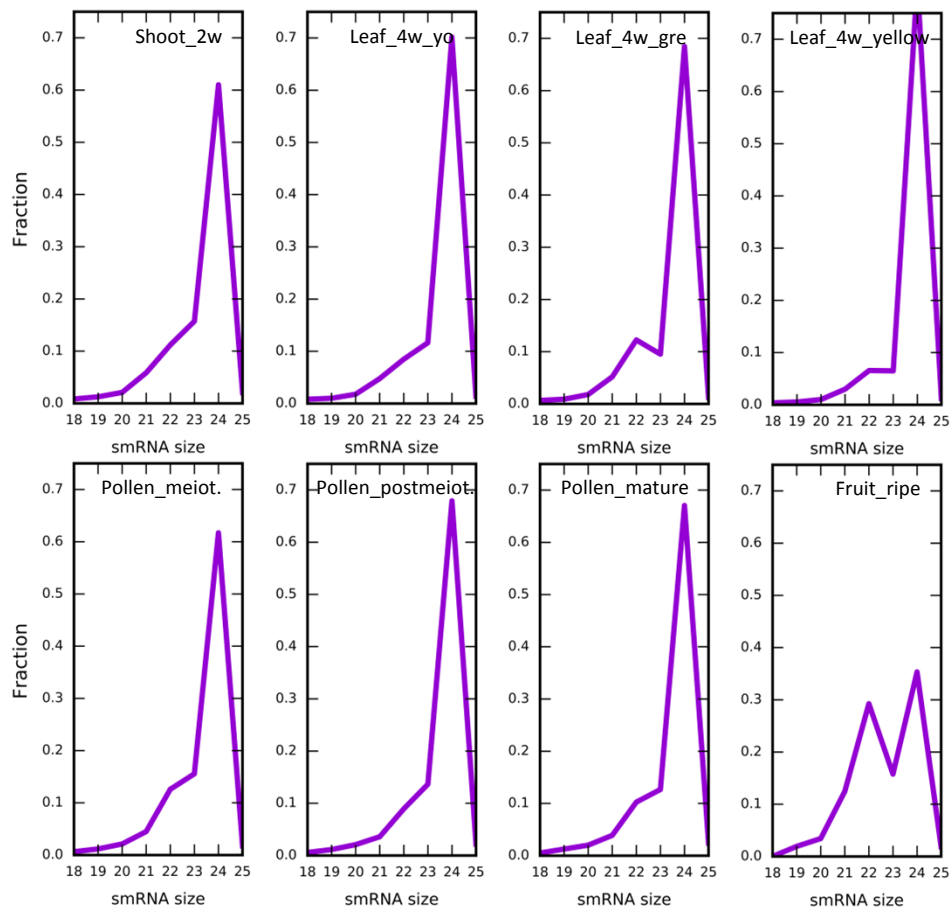
**A**

**non-truncated EPRVs (-TSAs) private smRNAs**



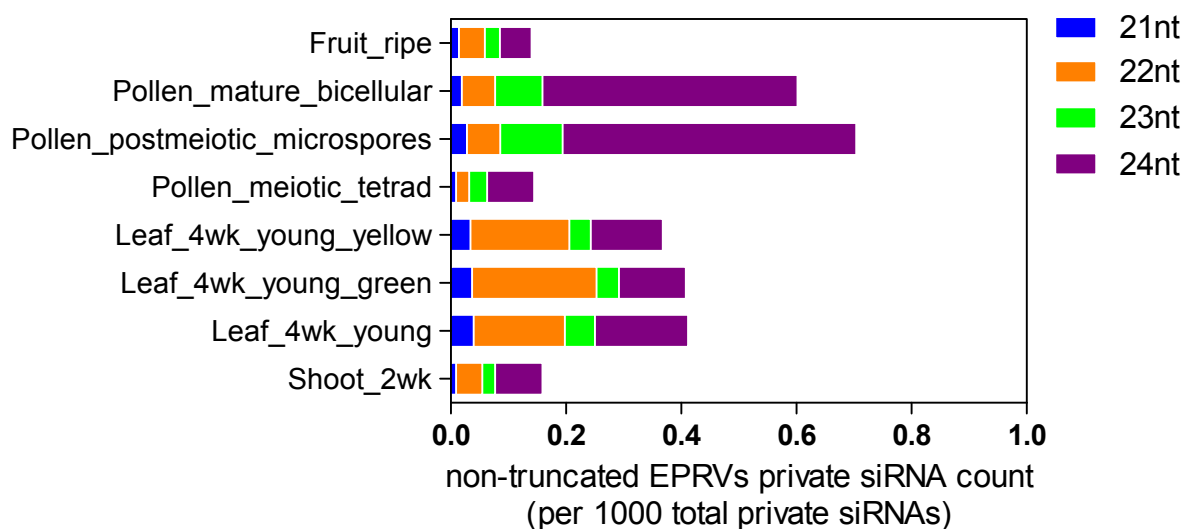
**B**

**Other pararetroviral sequences private smRNAs**

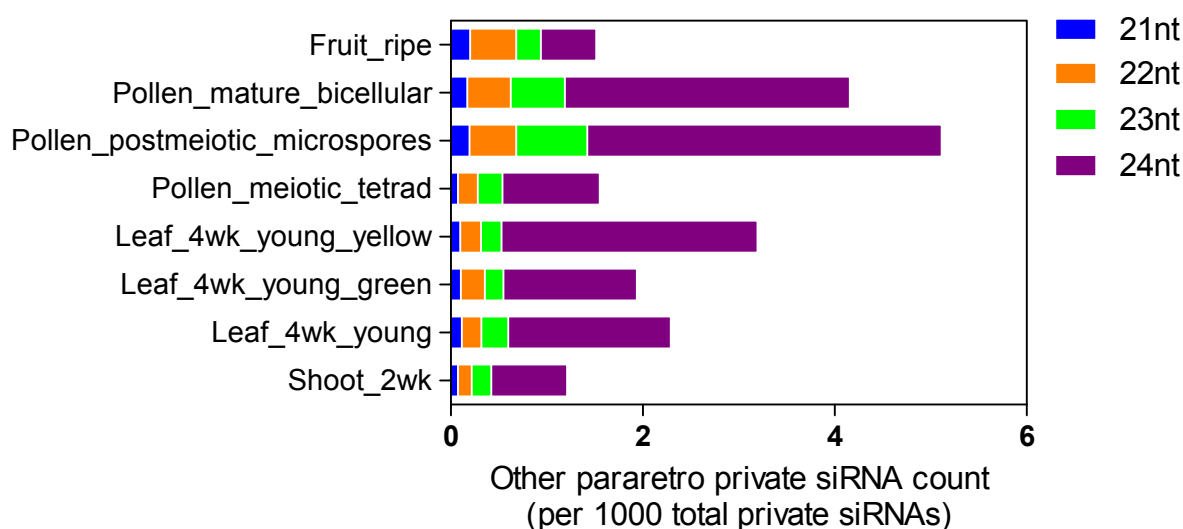


## Supplementary Figure S9

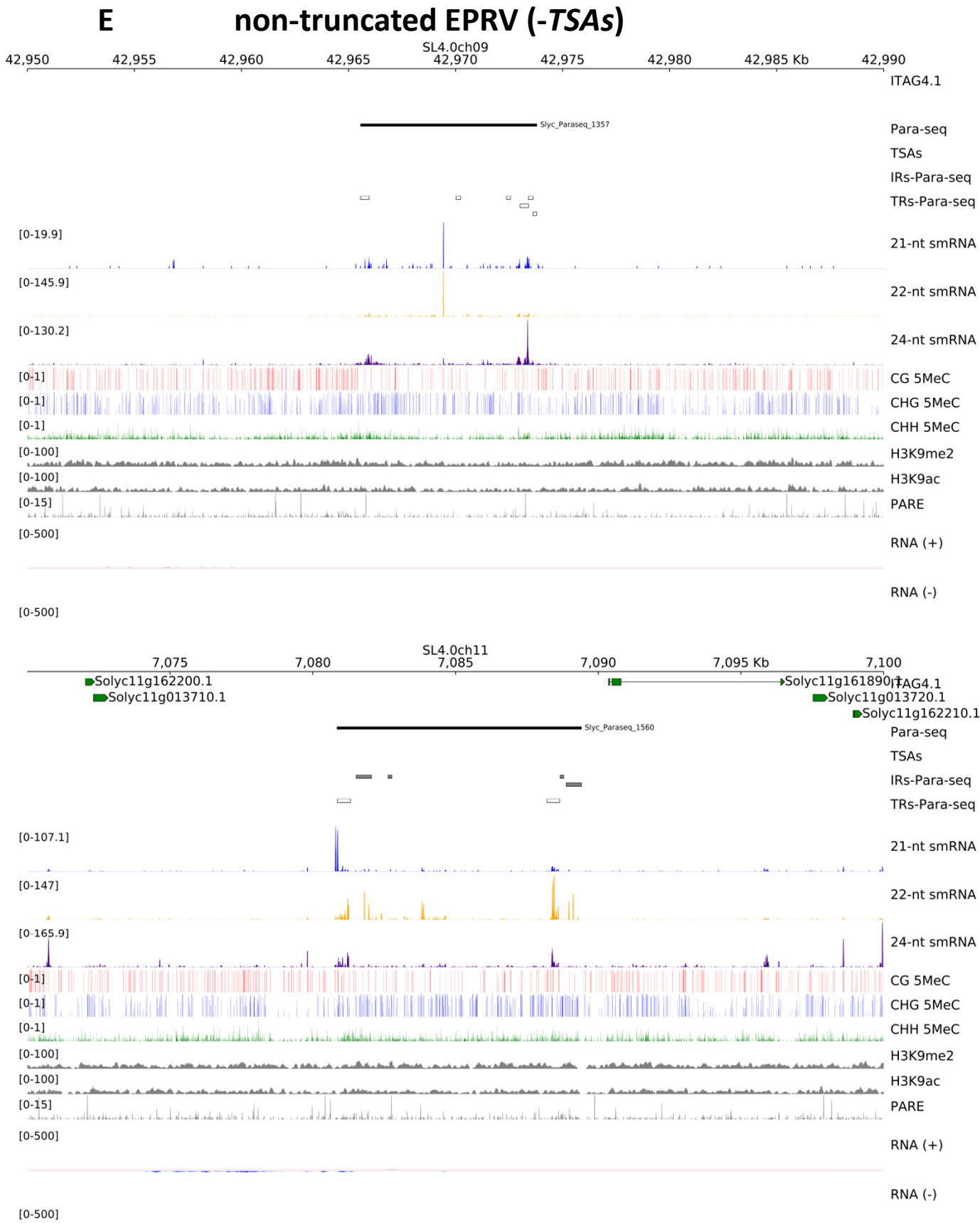
### C non-truncated EPRVs (-TSAs) private siRNAs



### D Other pararetroviral sequences' private siRNAs

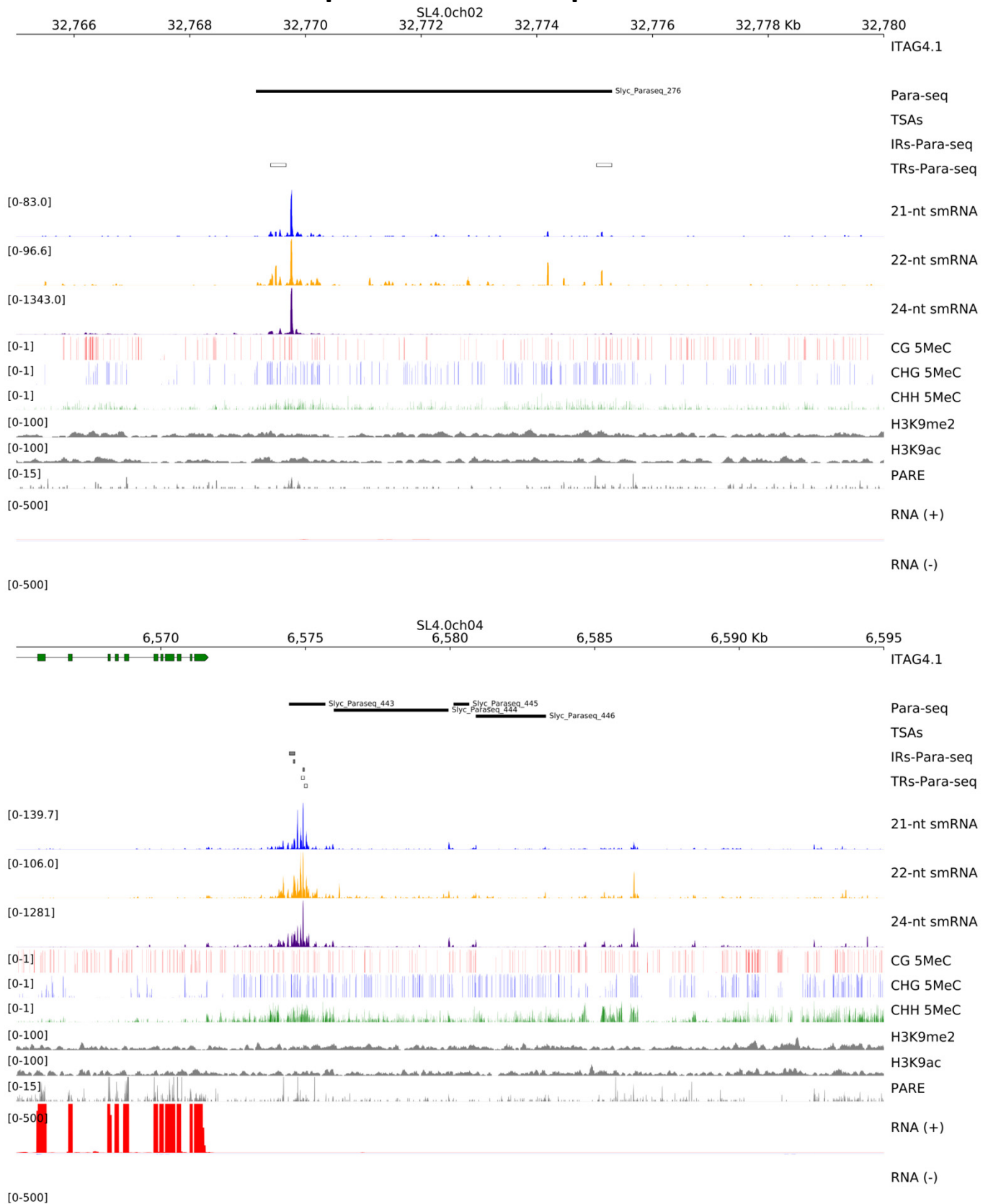


Supplementary Figure S9



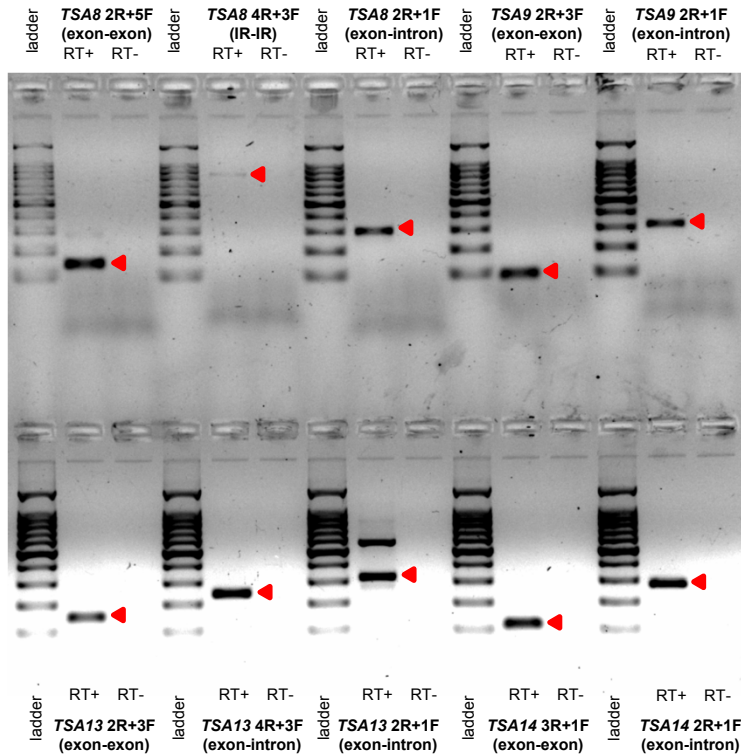
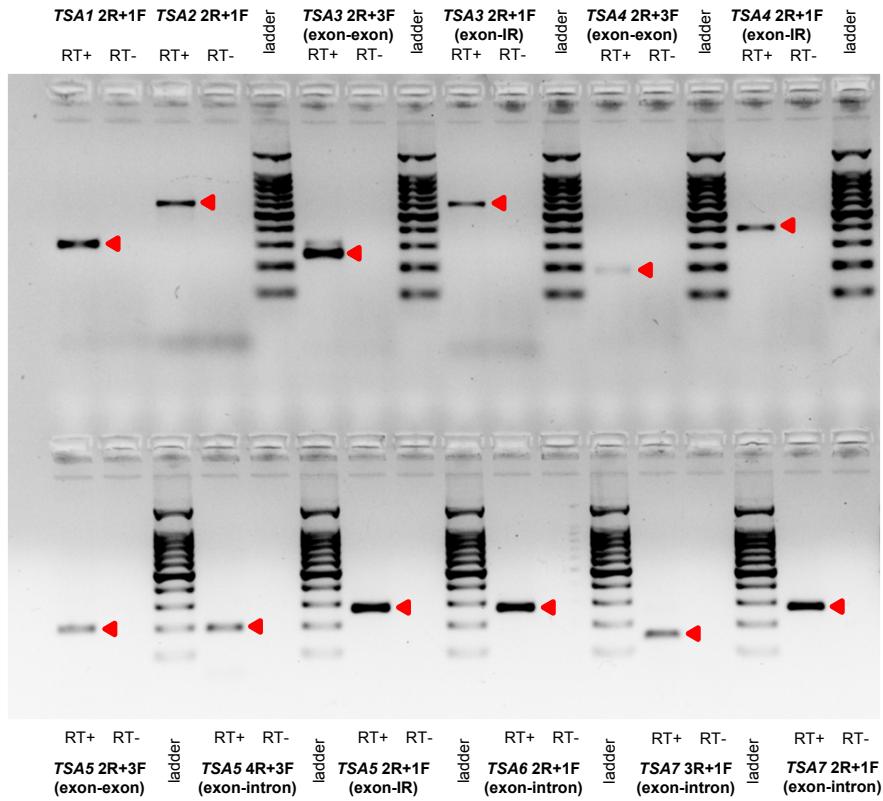
# Supplementary Figure S9

## F Other pararetroviral sequences

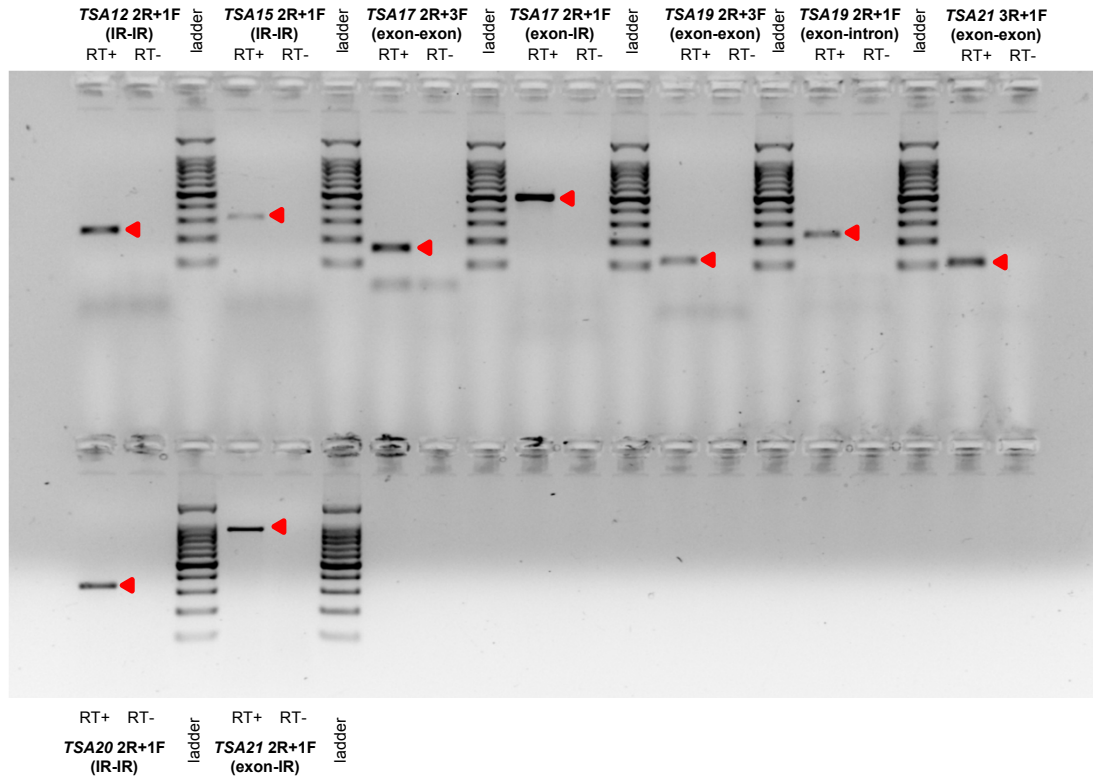


**Supplemental Figure S9:** siRNA profiles mapping to non-TSA coordinates. (A) Private siRNA size profiles across diverse tissues examined in non-truncated EPRVs (-TSAs) (i.e. not belonging to any TSA); and (B) other pararetroviral-related sequences not included as TSAs or non-truncated EPRVs (thus comprising historical remnant sequences). (C and D) Private siRNA count distinguished by size (21-nt = blue, 22-nt = orange, and 24-nt = indigo) across eight tissues mapping to non-TSA coordinates. (E and F) Genome-browsers around exemplary non-TSA coordinates. Data visualized as in Figure 3A.

# Supplementary Figure S10



## Supplementary Figure S10



**Supplemental Figure S10:** RT-PCR data from exemplified TSAs. Evidence for splicing can be inferred for genic TSAs in products from *TSA4*, *TSA5*, *TSA7*, *TSA8*, *TSA10*, *TSA13*, *TSA14*, *TSA17* and *TSA21*; and for intergenic TSAs in products from *TSA3*, *TSA9* and *TSA19*. Expected fragment sizes for each primer pair and primer sequence in Supplemental Table S8, whereas primers positions can be visualized in genome-browsers from Supplemental Figure S8. A second higher sized band in *TSA13* 2R+1F suggests an event of alternative splicing. Exon-exon = splicing product; exon-intron and exon-IR (inverted-repeat) = non-splicing product; IR-IR or no tag = confirmation of expression around or within TSA area; Ladder = 1000 bp ladder; RT+ = reaction with reverse-transcriptase; RT- = negative control reaction without addition of reverse-transcriptase.

# Supplementary Figure S11

IRS	5' IR size (bp)	3' IR size (bp)	% Identity between IRs	Assembled/template best hit (>50% query cover)	Sub-clade	Annotated Area/ORF
<b>TSA1a</b>	5007	4998	99.36	Slyc_Paraseq_900	<i>Solendovirus C1</i>	RT part., TAV, IGR, ORF1-4, CP, MP part.
<b>TSA2a</b>	834	833	99.64	Slyc_Paraseq_1592	<i>Florendovirus C1</i>	IGR, MP-CP-RT part.
<b>TSA2b</b>	3228	3240	98.95	Slyc_Paraseq_1592	<i>Florendovirus C1</i>	MP-CP-RT part., IGR
<b>TSA2c</b>	2116	2116	99.86	Slyc_Paraseq_48	<i>Florendovirus C2</i>	CP-RT part., ORF3, ORF4 part.
<b>TSA3a</b>	2788	2800	82.75	Slyc_Paraseq_889	<i>Solendovirus C1</i>	IGR, ORF1 part., ORF2 part., ORF3, ORF4, CP part.
<b>TSA3b</b>	387	388	94.83	Slyc_Paraseq_889	<i>Solendovirus C1</i>	TAV part.
<b>TSA4a</b>	6934	6949	99.71	Slyc_Paraseq_889	<i>Solendovirus C1</i>	IGR, ORF1 part., ORF3 part., ORF4, CP, MP, RT part., TAV part.
<b>TSA5a</b>	1630	1621	97.85	Spen_Paraseq_1326	<i>Caulimovirus/Soymovirus-related</i>	ORF5 part., IGR
<b>TSA6a</b>	712	715	98.32	Slyc_Paraseq_1592	<i>Florendovirus C1</i>	ORF2 part., IGR
<b>TSA7a</b>	333	333	96.40	-	-	-
<b>TSA8a</b>	1674	1679	96.68	Slyc_Paraseq_1918	<i>Caulimovirus/Soymovirus-related</i>	ORF5 part., IGR
<b>TSA9a</b>	1506	1514	96.06	Spen_Paraseq_1538	<i>Florendovirus C2</i>	ORF4 part., IGR
<b>TSA10a</b>	1311	1312	97.69	Spen_Paraseq_1538	<i>Florendovirus C2</i>	ORF4 part., IGR
<b>TSA11a</b>	1946	1982	93.01	Slyc_Paraseq_345	<i>Solendovirus C1</i>	TAV part., IGR
<b>TSA12a</b>	324	325	99.69	Slyc_Paraseq_345	<i>Solendovirus C1</i>	IGR
<b>TSA12b</b>	1769	1756	98.57	Slyc_Paraseq_345	<i>Solendovirus C1</i>	IGR, ORF1, ORF2, ORF3, ORF4, CP part.
<b>TSA13a</b>	562	552	95.90	Slyc_Paraseq_1918	<i>Caulimovirus/Soymovirus-related</i>	ORF5 part., IGR
<b>TSA15a</b>	1392	1392	94.69	Slyc_Paraseq_889	<i>Solendovirus C1</i>	Non-coding
<b>TSA15b</b>	1542	1542	96.11	Spen_Paraseq_947	<i>Solendovirus C1</i>	IGR, ORF1, ORF2, ORF3, ORF4, CP part.
<b>TSA15c</b>	1179	1179	95.41	Slyc_Paraseq_889	<i>Solendovirus C1</i>	Non-coding
<b>TSA15d</b>	1070	1073	74.95	Slyc_Paraseq_900	<i>Solendovirus C1</i>	TAV part.
<b>TSA16a</b>	143	143	97.18	Slyc_Paraseq_345	<i>Solendovirus C1</i>	IGR
<b>TSA16b</b>	769	770	99.09	Spen_Paraseq_947	<i>Solendovirus C1</i>	CP part., MP part.
<b>TSA17a</b>	2236	2234	99.24	Slyc_Paraseq_889	<i>Solendovirus C1</i>	RT part., TAV, IGR
<b>TSA17b</b>	4176	4176	99.90	Slyc_Paraseq_889	<i>Solendovirus C1</i>	MP part., RT, TAV, IGR
<b>TSA18a</b>	3720	3730	98.15	Spen_Paraseq_1538	<i>Florendovirus C2</i>	CP-RT part., ORF3, ORF4, IGR
<b>TSA19a</b>	3668	3880	98.71	Slyc_Paraseq_900	<i>Solendovirus C1</i>	MP part., RT part., TAV, IGR
<b>TSA20a</b>	644	631	94.40	Slyc_Paraseq_1918	<i>Caulimovirus/Soymovirus-related</i>	IGR
<b>TSA20b</b>	3015	3064	82.23	Slyc_Paraseq_1918	<i>Caulimovirus/Soymovirus-related</i>	RT part., ORF5, IGR
<b>TSA20c</b>	1453	1457	96.43	Slyc_Paraseq_1918	<i>Caulimovirus/Soymovirus-related</i>	ORF5 part., IGR
<b>TSA21a</b>	1718	1734	96.09	Slyc_Paraseq_1918	<i>Caulimovirus/Soymovirus-related</i>	ORF5 part., IGR
<b>TSA22a</b>	1908	1825	98.52	Slyc_Paraseq_48	<i>Florendovirus C2</i>	ORF4 part., IGR

**Supplemental Figure S11:** *In silico* analysis of TSAs' relevant inverted-repeat sequences. CP = capsid protein; MP = movement protein; RT = reverse-transcriptase; TAV = transactivator/viroplasm; ORFX = unknown; Part. = partial; IGR = inter-genic region.

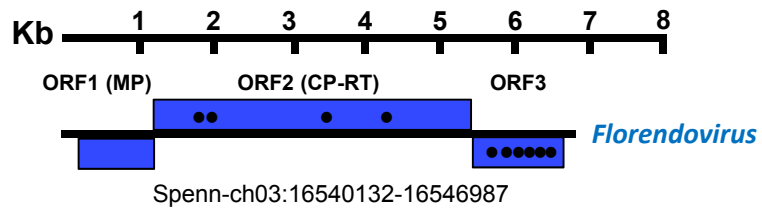
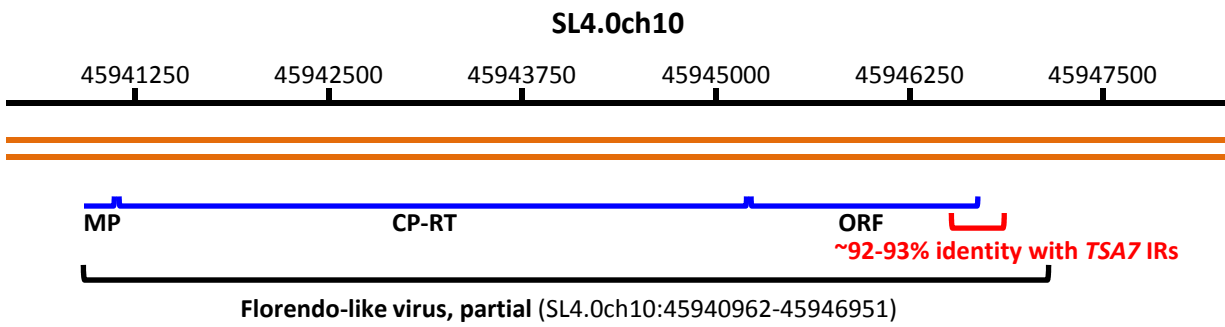
# Supplementary Figure S12

	Solendovirus							Florendovirus C1				Florendovirus C2				Caulimovirus/Soymovirus-related			non-redundant smRNA hits from considered targets (% TSA-mapping total)	List non-truncated EPRVs	
	Slyc_Paraseq_345	Slyc_Paraseq_889	Slyc_Paraseq_900	Spen_Paraseq_947	Stub_Paraseq_5519†	Smel_Paraseq_2976	Tobacco vein clearing virus	Slyc_Paraseq_1592	Spen_Paraseq_330	Stub_Paraseq_2284	Smel_Paraseq_6042*	Slyc_Paraseq_48	Spen_Paraseq_1538	Stub_Paraseq_3500†	Smel_Paraseq_1563	Slyc_Paraseq_1918*	Spen_Paraseq_1326	Stub_Paraseq_4874‡		non-redundant smRNA hits (% non-truncated EPRVs mapping total)	Number of targeted non-truncated EPRVs (-TSAs)
TSA1	8.4	5.9	25.1	2.5	0.8	0.1	0.4	-	-	-	-	-	-	-	-	-	-	-	30.1	29.30	114
TSA2	-	-	-	-	-	<0.1	-	9.2	3.1	0.1	-	0.9	0.6	-	0.1	-	-	-	10.7	18.46	16
TSA3	25.8	18.3	21.4	11.4	0.3	-	0.4	-	-	-	-	-	-	-	-	-	-	-	39.7	67.94	114
TSA4	1.9	4.2	2.4	4.7	0.3	0.2	0.3	-	-	-	-	-	-	-	-	-	-	-	9.2	28.48	114
TSA5	-	-	-	-	-	-	-	<0.1	-	-	-	-	-	-	-	-	0.8	1.0	1.6	<0.01	1
TSA6	-	-	-	-	-	-	-	<0.1	<0.1	-	-	-	-	-	-	-	-	-	0.1	21.75	5
TSA7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
TSA8	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.6	2.5	<0.1	3.8	-
TSA9	-	-	-	-	-	-	-	-	-	-	-	0.1	0.1	<0.1	2.6	-	-	-	2.8	0.31	7
TSA10	-	-	-	-	-	-	-	-	-	-	-	<0.1	0.5	-	<0.1	-	-	-	0.5	1.56	11
TSA11	<0.1	-	-	-	-	<0.1	-	-	-	-	-	-	-	-	-	-	-	-	<0.1	0.90	10
TSA12	13.5	7.5	12.3	4.6	<0.1	0.1	0.3	-	-	-	-	-	-	-	-	-	-	-	19.0	61.64	114
TSA13	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.9	4.3	<0.1	4.4	-
TSA14	0.5	2.0	2.4	4.5	1.0	<0.1	0.2	-	-	-	-	-	-	-	-	-	-	-	8.1	51.02	114
TSA15	2.5	3.4	3.0	2.1	0.1	<0.1	0.3	-	-	-	-	-	-	-	-	-	-	-	6.0	41.31	114
TSA16	1.6	3.2	1.1	5.2	0.3	0.5	0.8	-	-	-	-	-	-	-	-	-	-	-	7.1	16.49	114
TSA17	0.7	2.2	0.7	3.6	0.3	0.5	1.7	-	-	-	-	-	-	-	-	-	-	-	6.0	13.40	114
TSA18	-	-	-	-	-	-	-	<0.1	-	-	-	2.9	0.9	-	0.3	-	-	-	3.8	3.8	10
TSA19	1.2	3.3	2.8	4.9	0.5	0.2	0.4	-	-	-	-	-	-	-	-	-	-	-	8.5	44.78	114
TSA20	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	30.4	15.3	<0.1	32.7	-
TSA21	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	60.9	18.0	0.6	61.4	-
TSA22	-	-	-	-	-	-	-	-	-	-	-	2.5	0.2	-	0.1	-	-	-	2.7	2.38	7

**Supplemental Figure S12:** Analysis of TSAs' mapping 21-22-24-nt siRNAs complementary to *in silico* target templates. Pooled libraries were considered for perfect-matching and non-perfect-matching (only one allowed SNP in the last 10 bp of molecules) probes. For sub-clade targets, the fraction of complementary siRNAs is shown as a percentage of the total 21-22-24-nt forms mapping to the corresponding TSAs, whereas the non-redundant sum is shown in column #20.

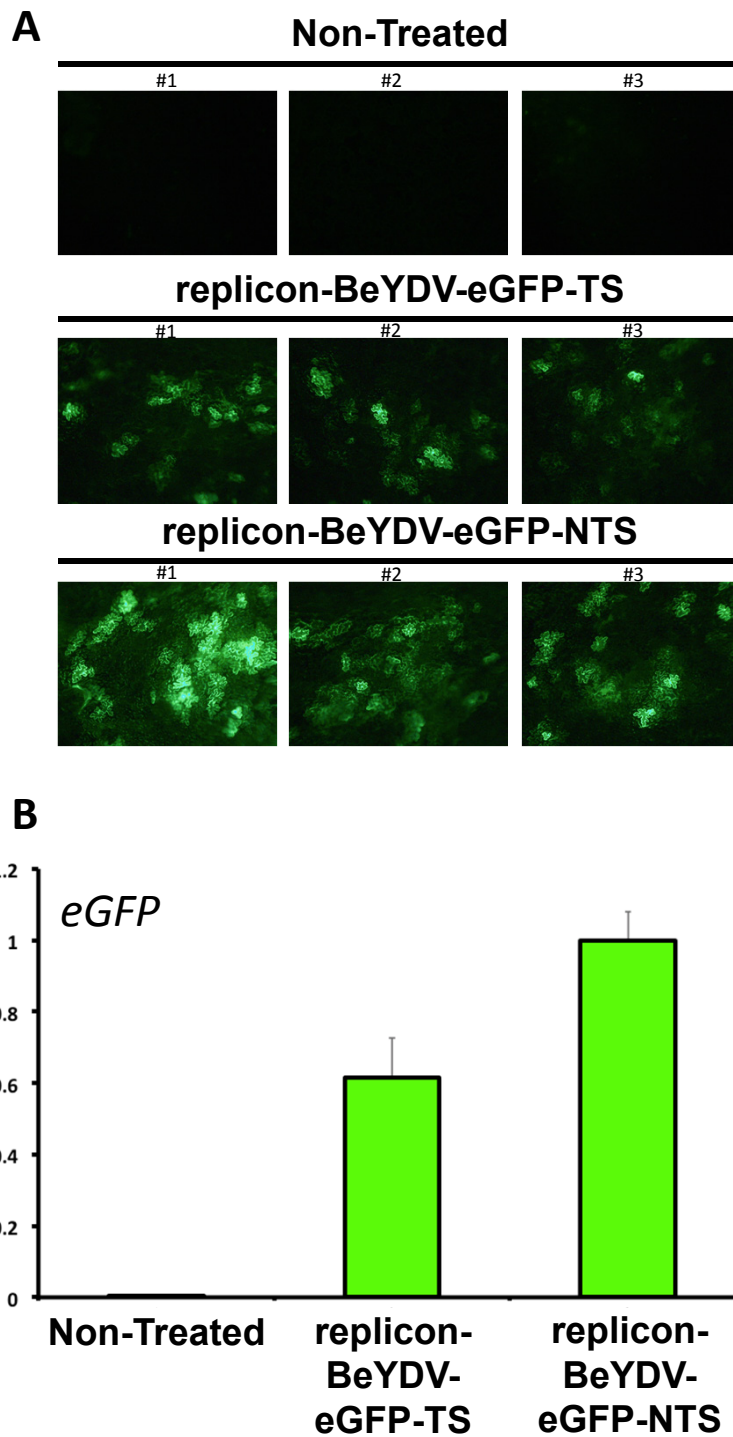


## Supplementary Figure S13



**Supplemental Figure S13:** Discovery of an elder *Florendovirus*-like EPRV bearing sequences matching *TSA7*. Note that this SL4.0ch10:45940962-45946951 sequence (top) is too derived for correct assembly, but a more complete candidate was found in *S. pennellii* (bottom, Spenn-ch03:16540132-16546987 not listed among the non-truncated EPRVs of Supplemental Table S2), confirming these are *Florendovirus* with three ORF configuration. Boxes represent ORFs, with black dots indicating positions with premature stop codons.

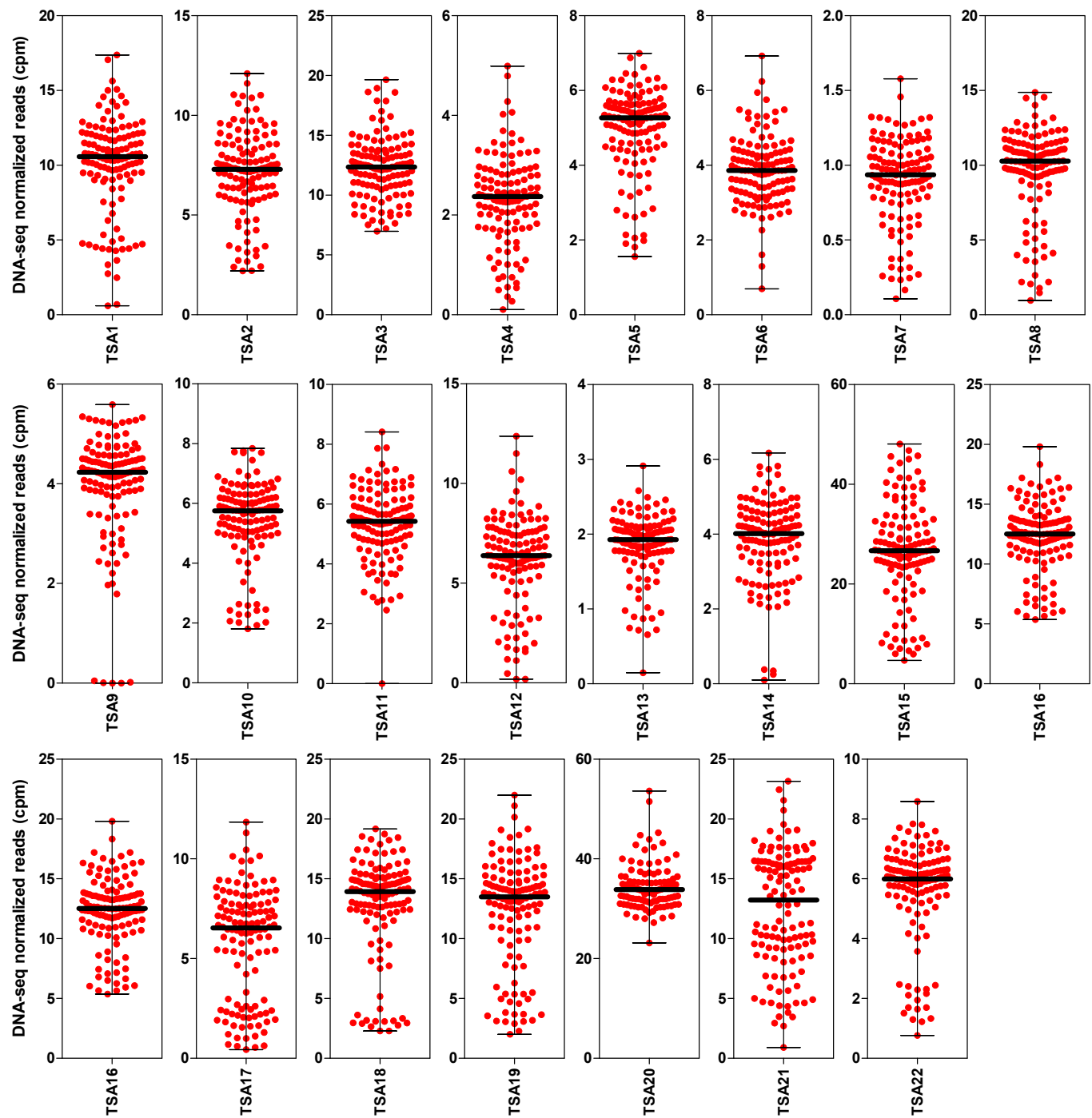
## Supplementary Figure S14



**Supplemental Figure S14:** Reporter expression in replicon-BeYDV-eGFP-TS and replicon-BeYDV-eGFP-NTS infiltrations. **(A)** *S. lycopersicum* cotyledons three days after infiltration under UV light revealing eGFP fluorescence (microscope picture). Three independent plants were infiltrated with the indicated constructs; non-treated plants were also included as negative controls. **(B)** Precise *eGFP* transcript level as measured by RT-qPCR in tissues from the previous cotyledons; each bar represents the average of those three replicated plants (error bar = 1 standard deviation). *eGFP* transcript level was normalized to Actin as housekeeping gene, and then relativized to replicon-BeYDV-eGFP-NTS (equal to 1). Expression differences were significant at  $p < 0.01$  (One-way ANOVA with post-hoc Tukey Test).

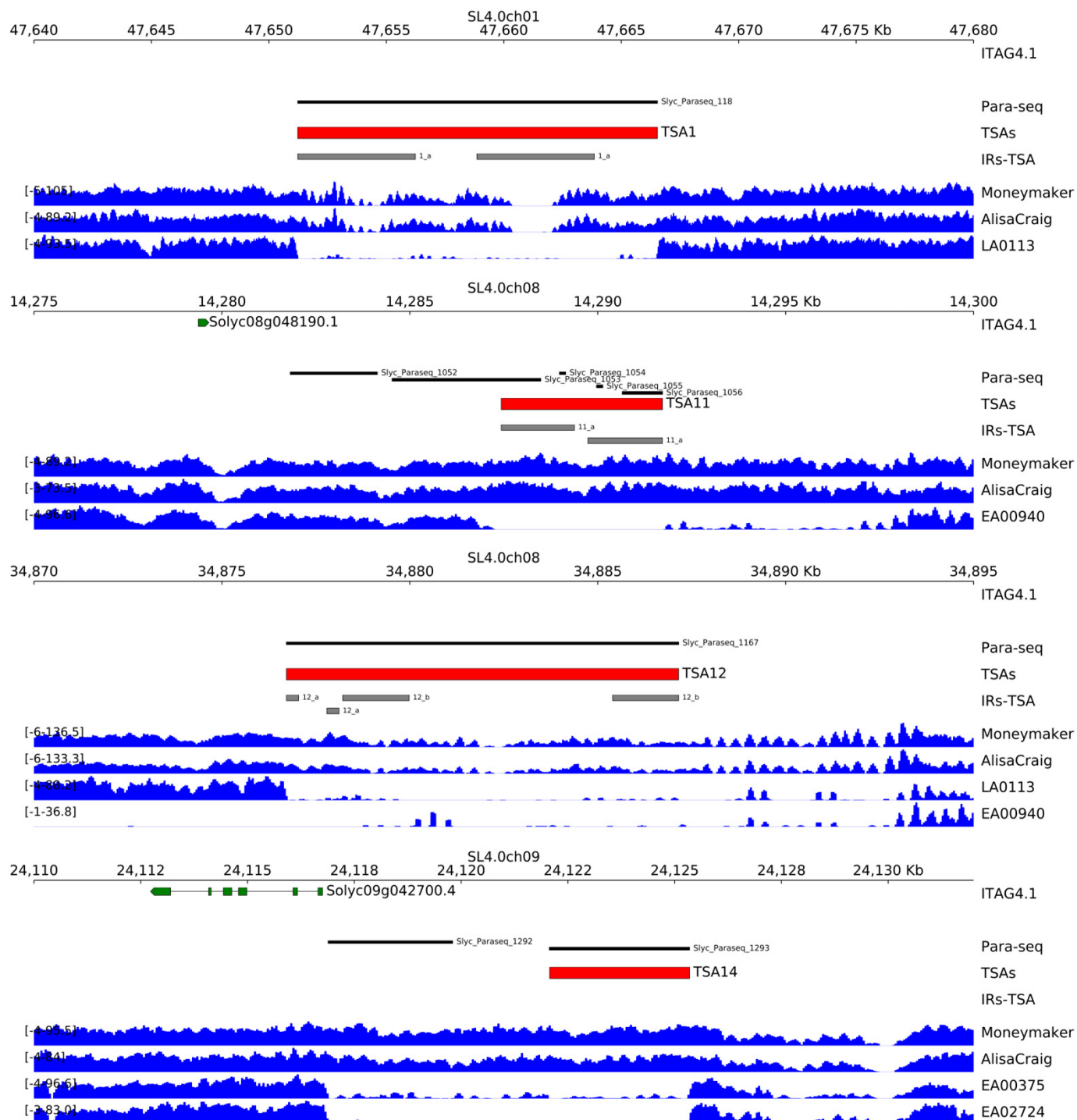
# Supplementary Figure S15

A



# Supplementary Figure S15

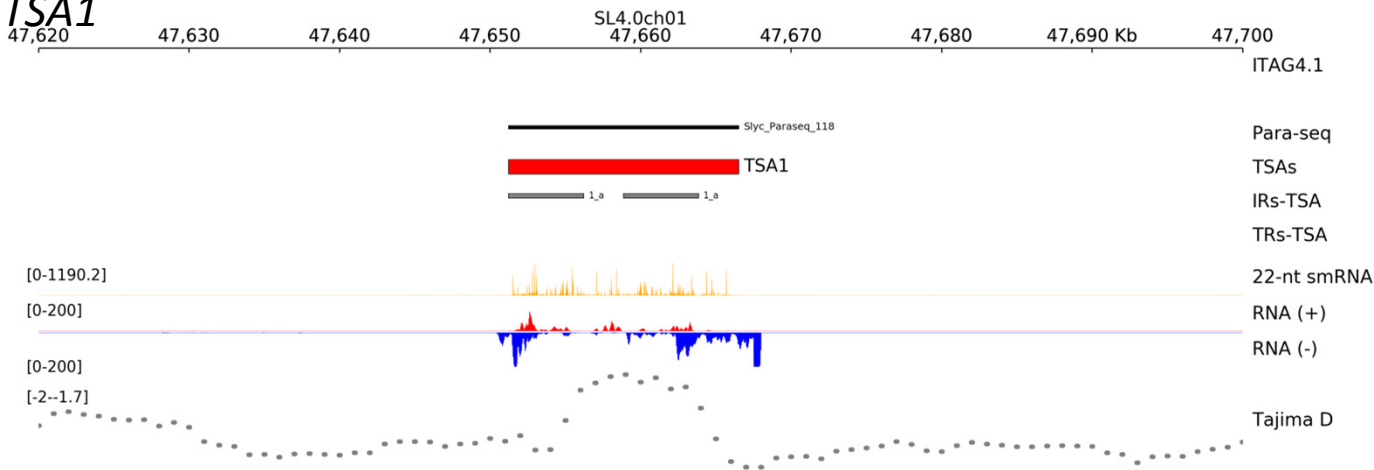
**B**



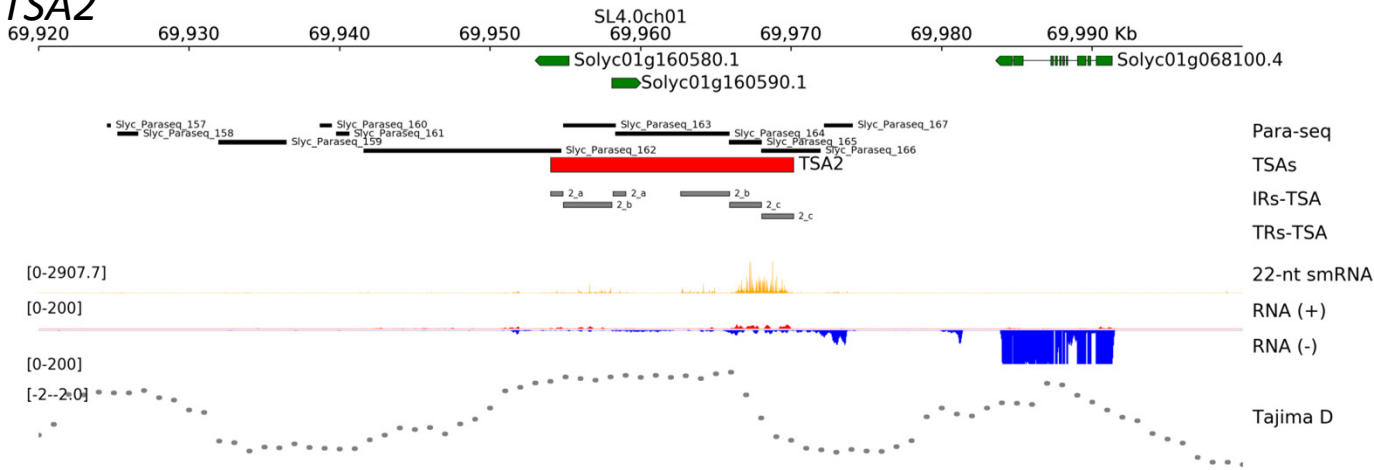
**Supplemental Figure S15:** (A) Counts of private DNA-seq TSA-mapping reads across 124 *S. lycopersicum* accessions bearing at least x5 private reads genome coverage, normalized by library size (expressed as count-per-million, cpm). Median and range were added to the scattered dot plot. (B) Genome-browser for private TSA-mapping reads (blue) in exemplified lower-end outlier accessions. Recognized pararetroviral sequences (Para-seq, black) and inverted-repeats (IRs-TSA, grey) are shown. Moneymaker and AlisaCraig accessions are presented as positive controls.

# Supplementary Figure S16

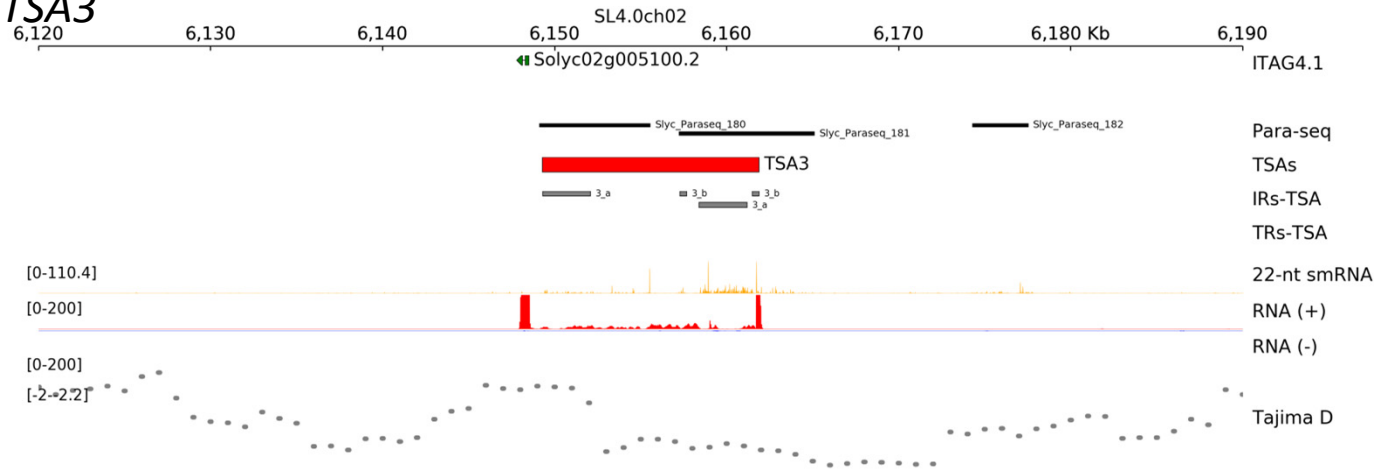
## TSA1



## TSA2

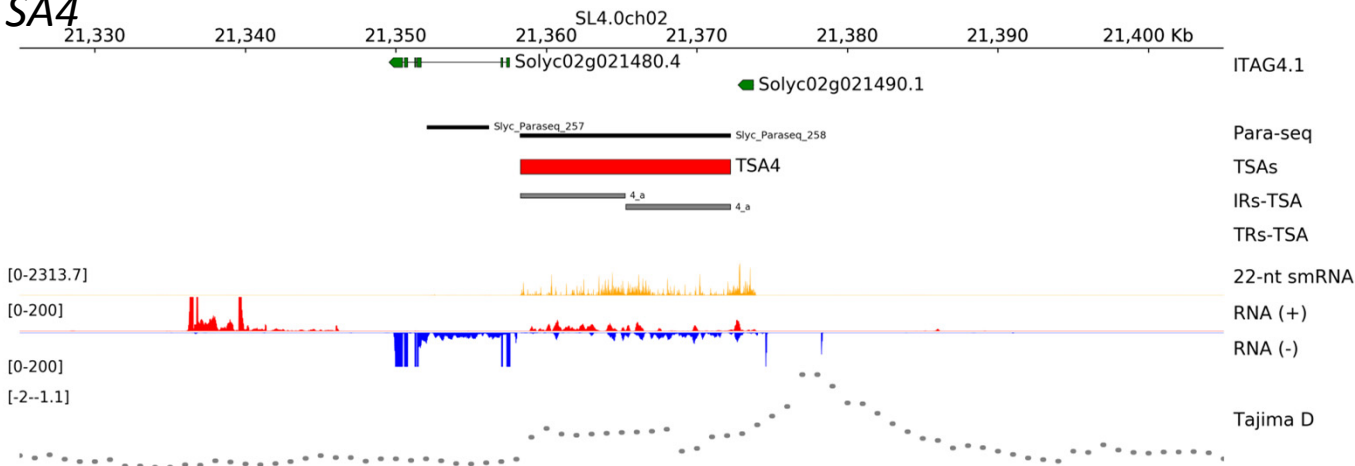


## TSA3

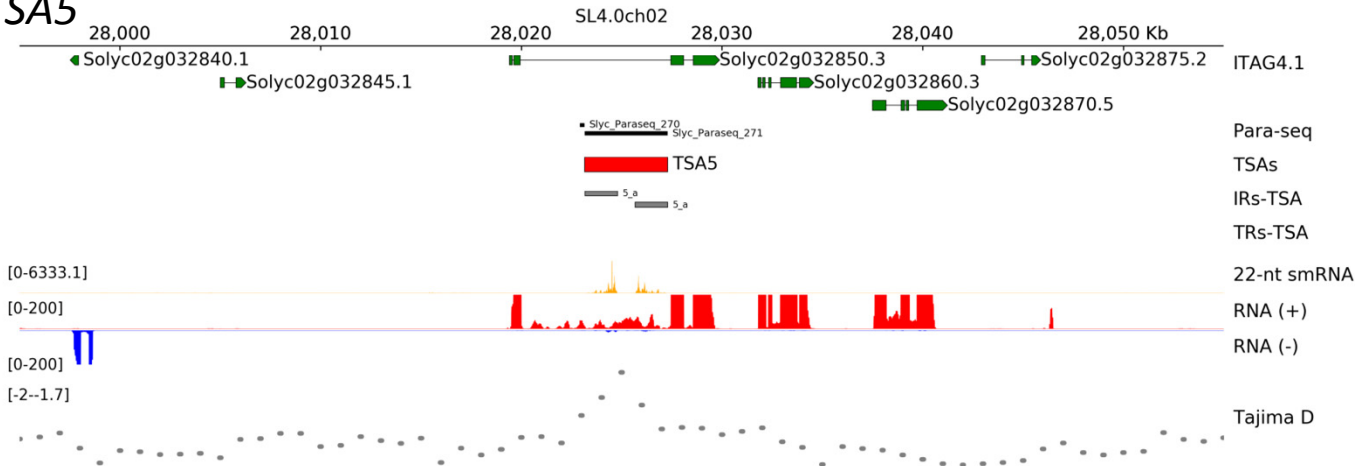


Supplementary Figure S16

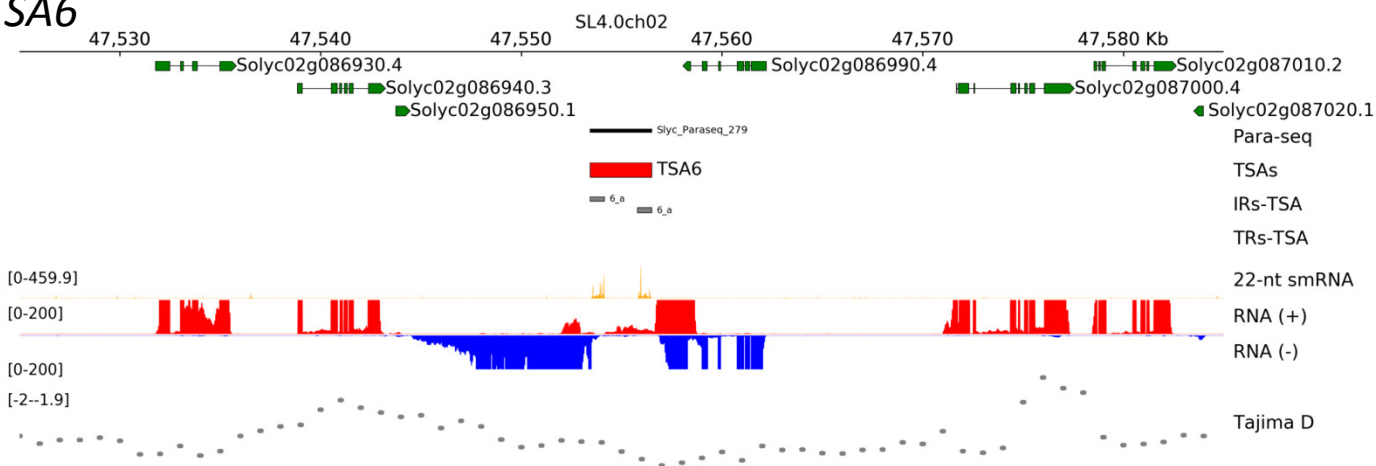
TSA4



TSA5

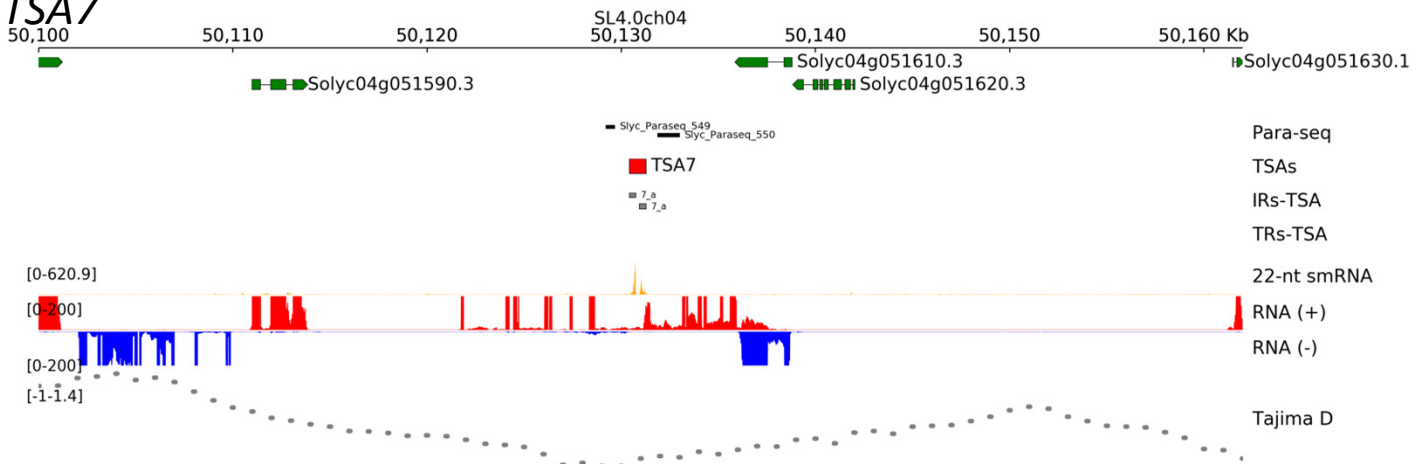


TSA6

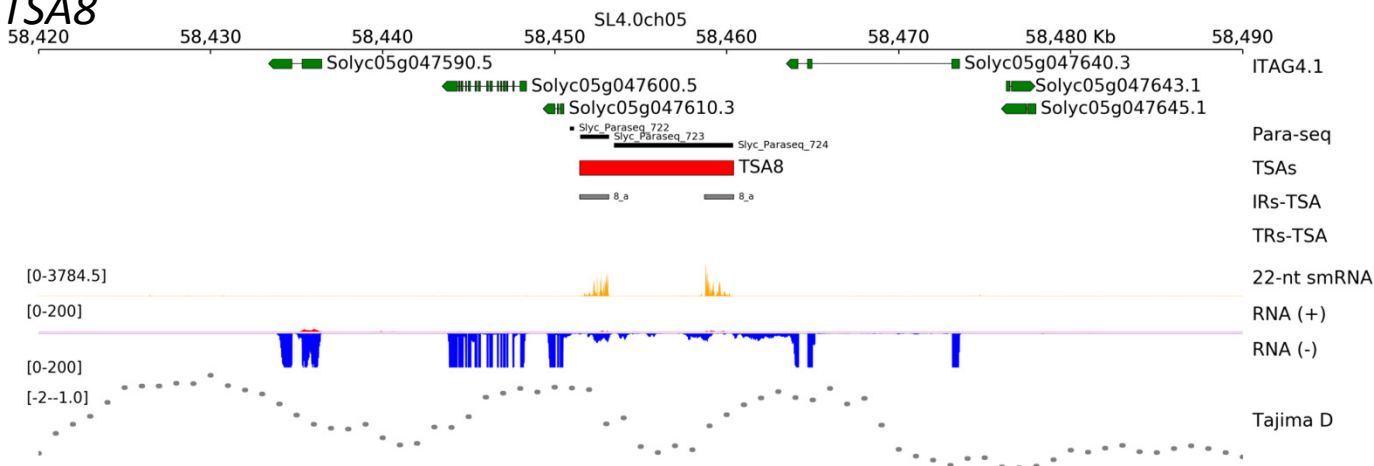


# Supplementary Figure S16

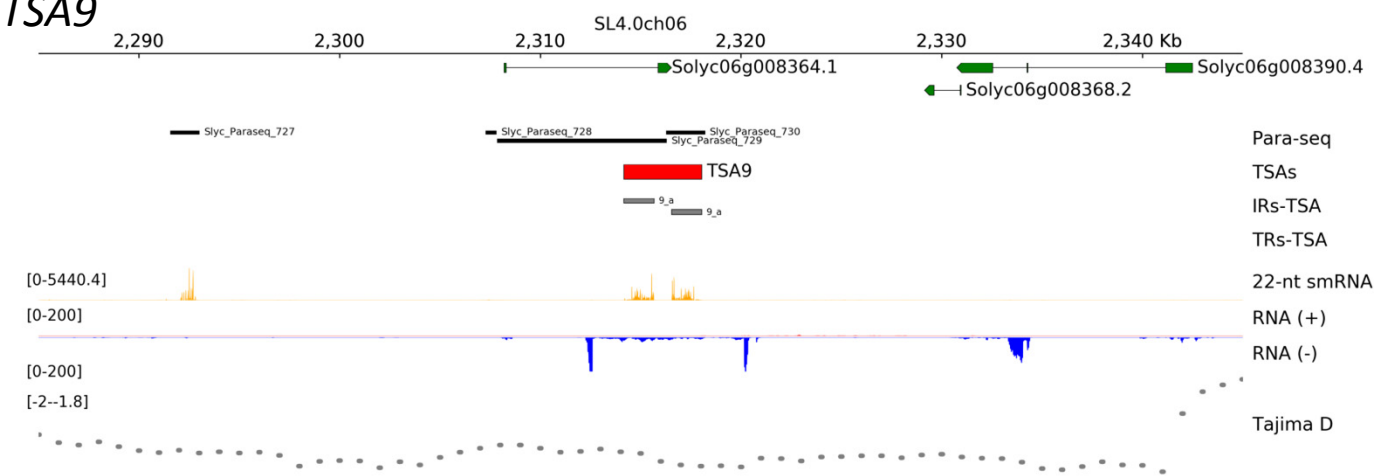
## TSA7



## TSA8

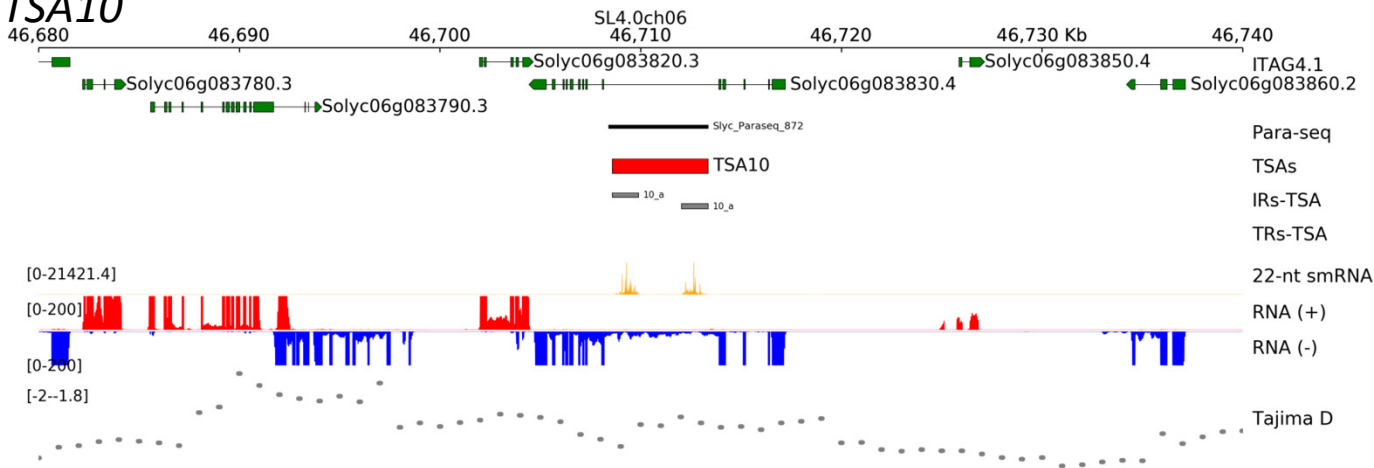


## TSA9

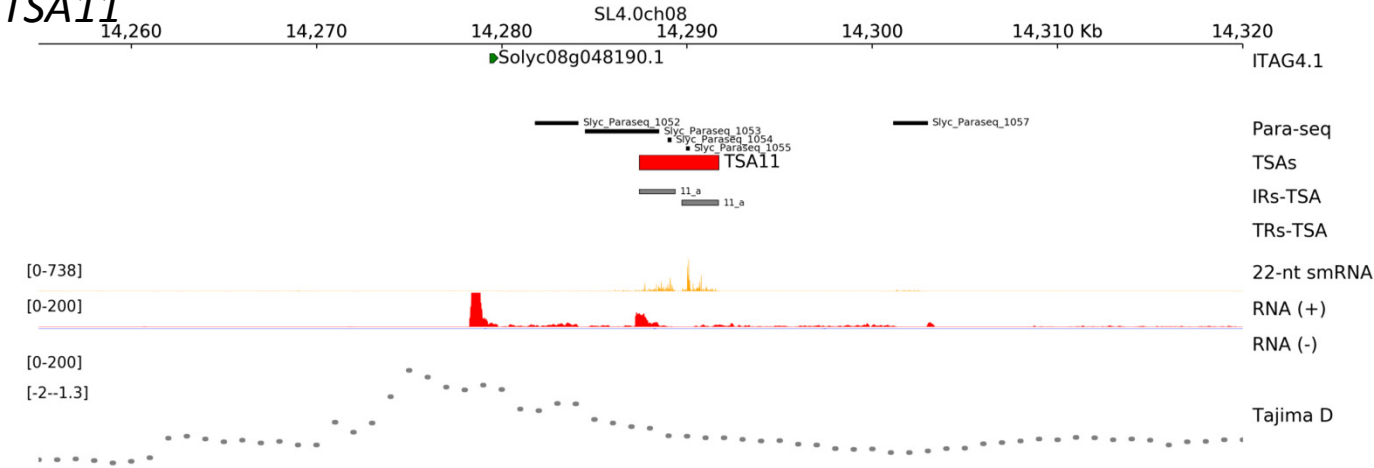


# Supplementary Figure S16

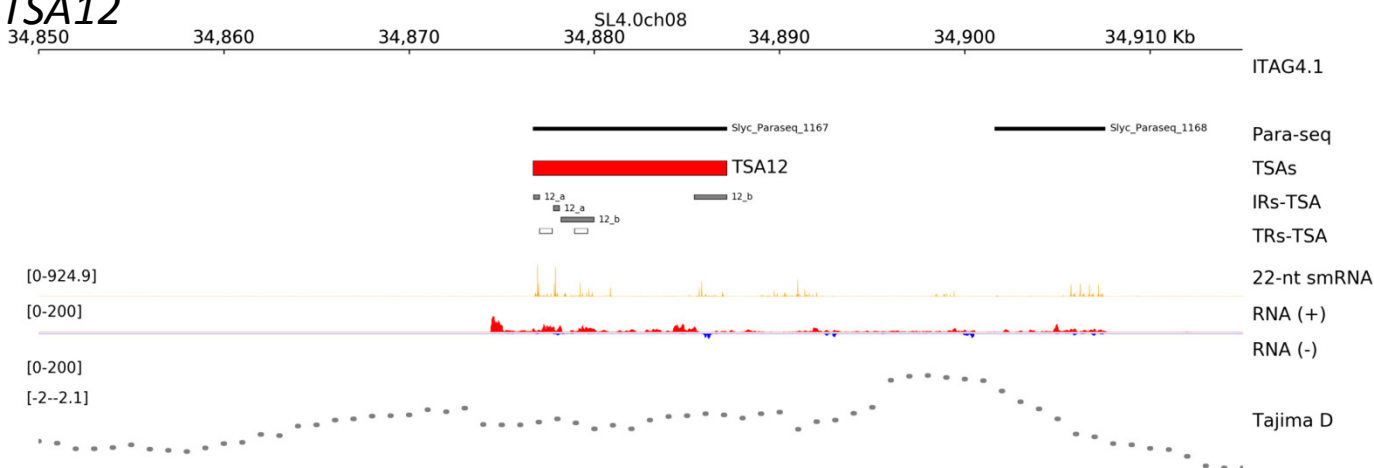
## TSA10



## TSA11



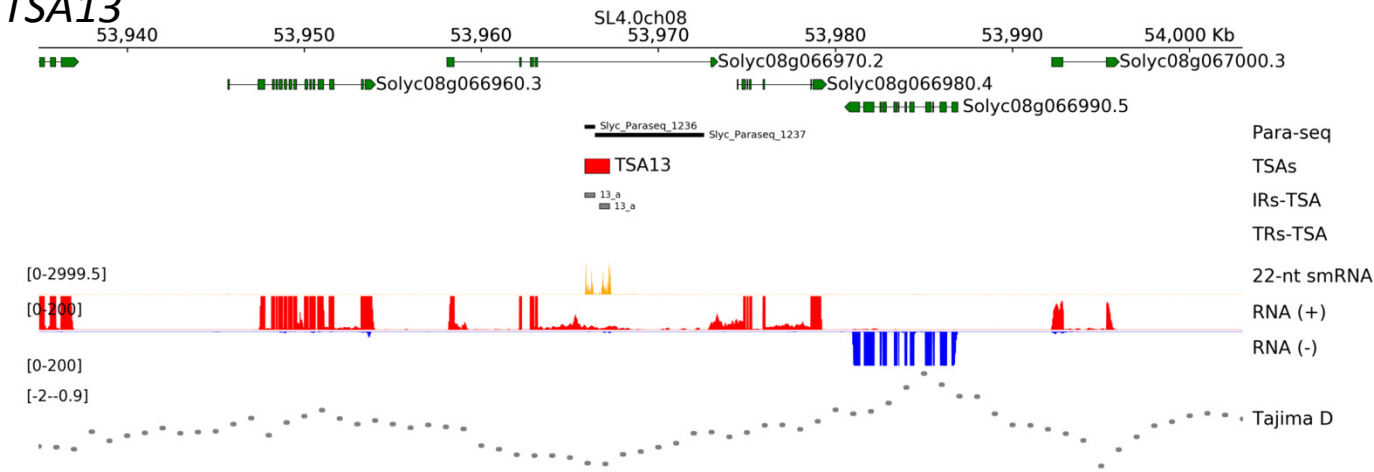
## TSA12



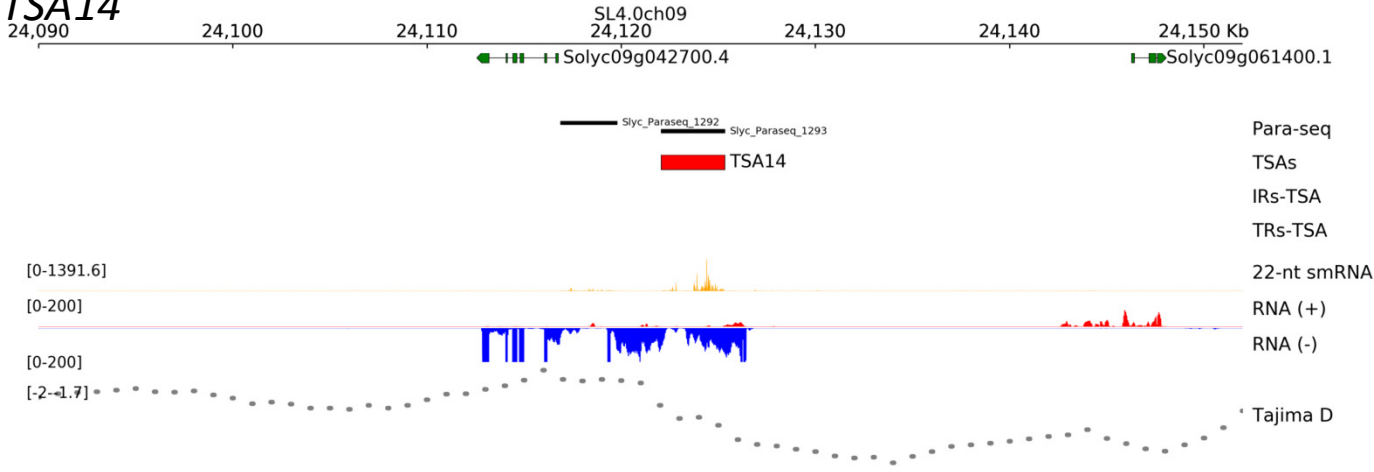


Supplementary Figure S16

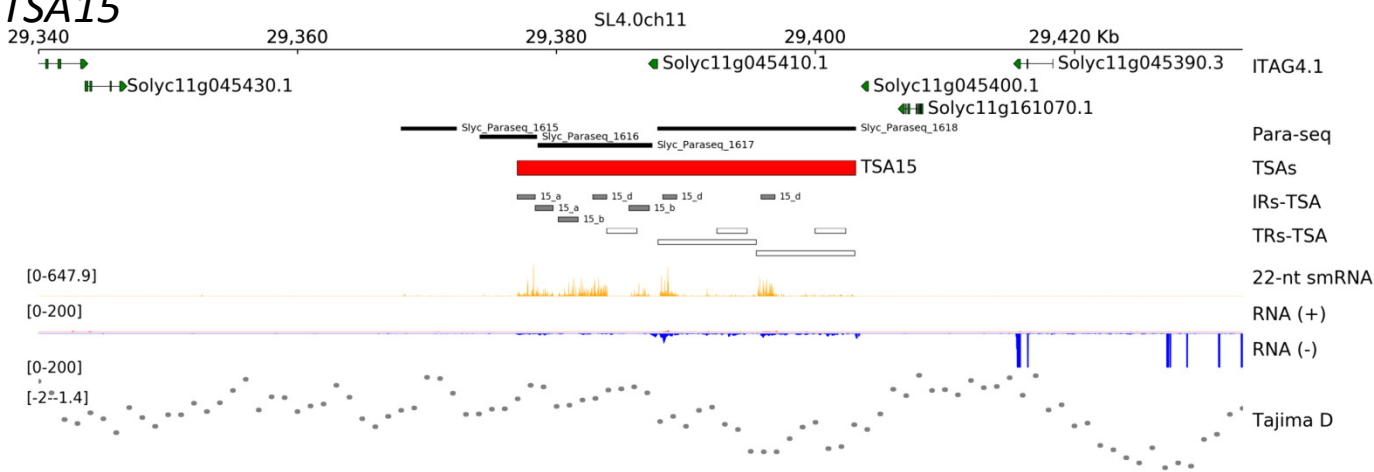
TSA13



TSA14

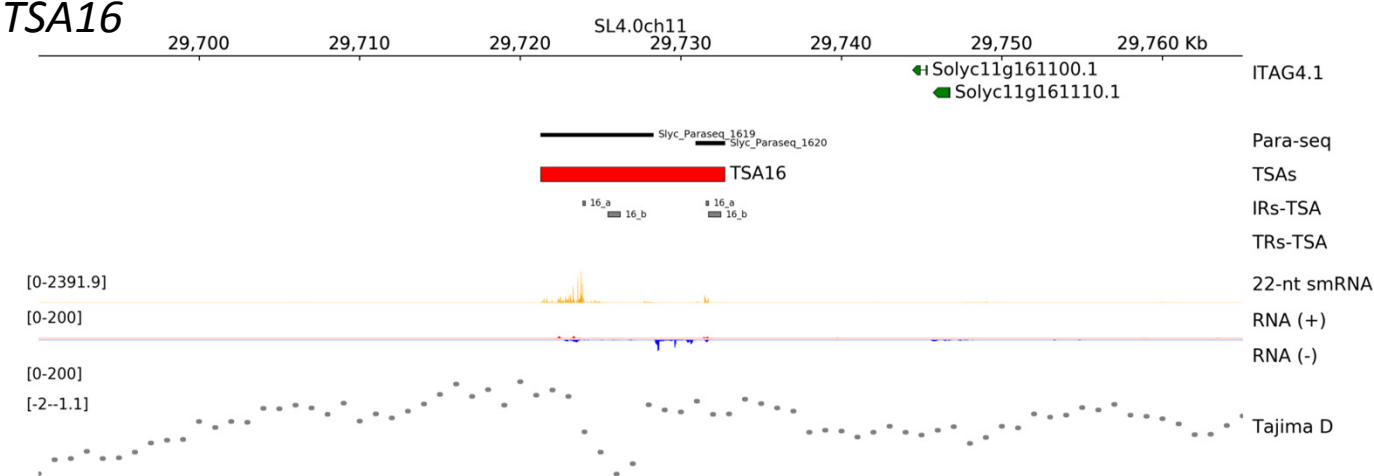


TSA15

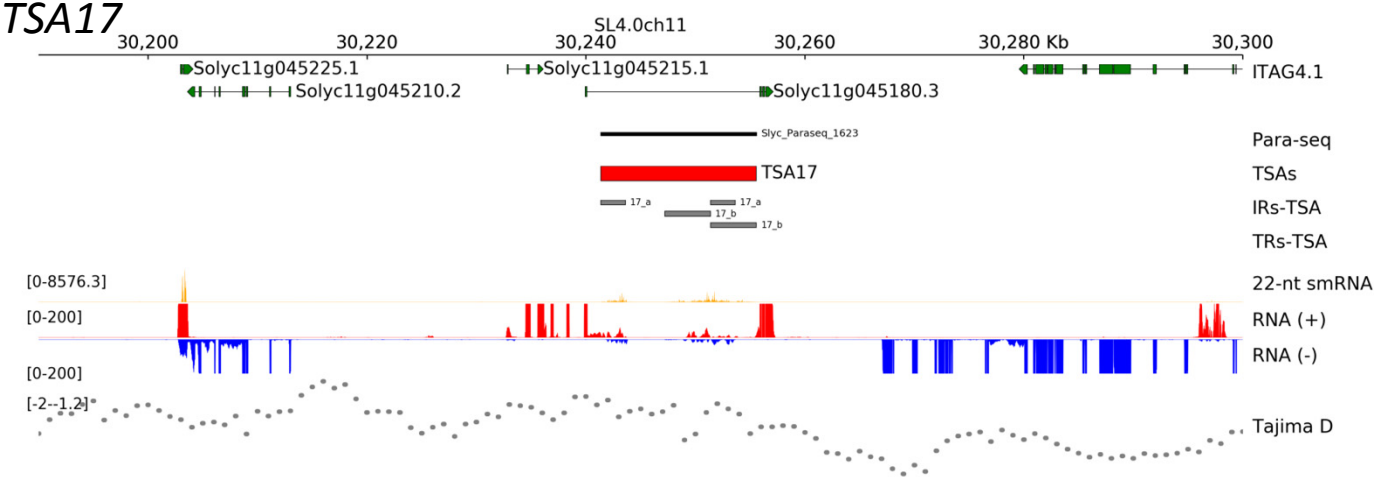


# Supplementary Figure S16

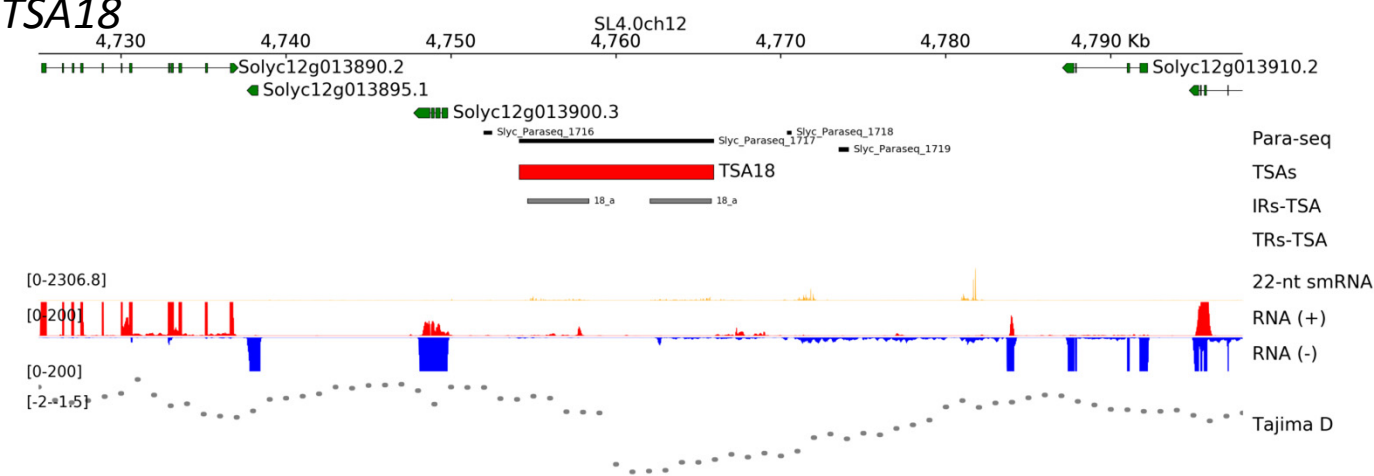
## TSA16



## TSA17

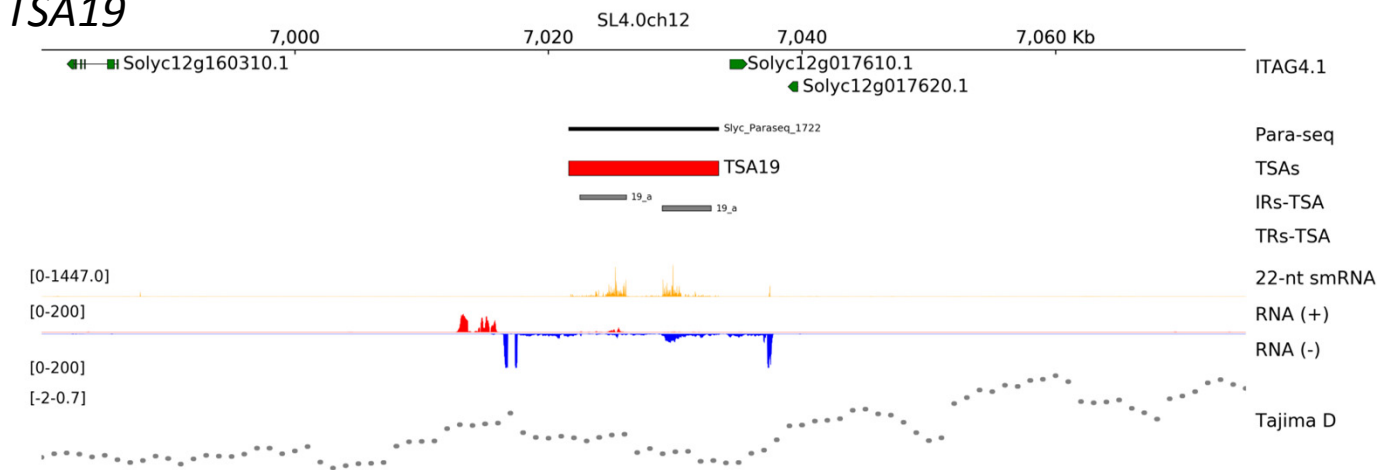


## TSA18

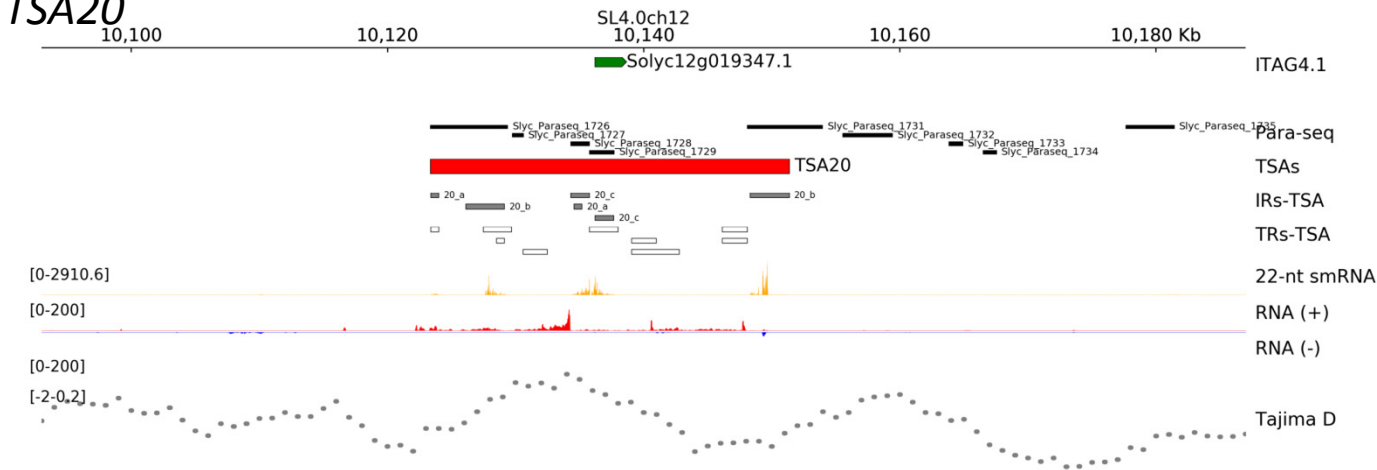


# Supplementary Figure S16

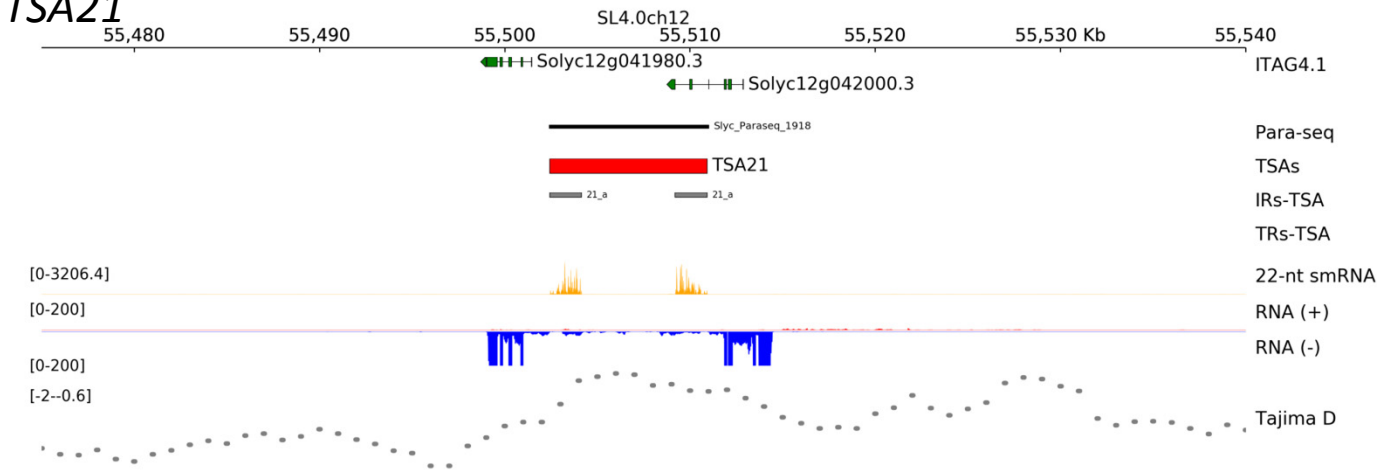
## TSA19



## TSA20

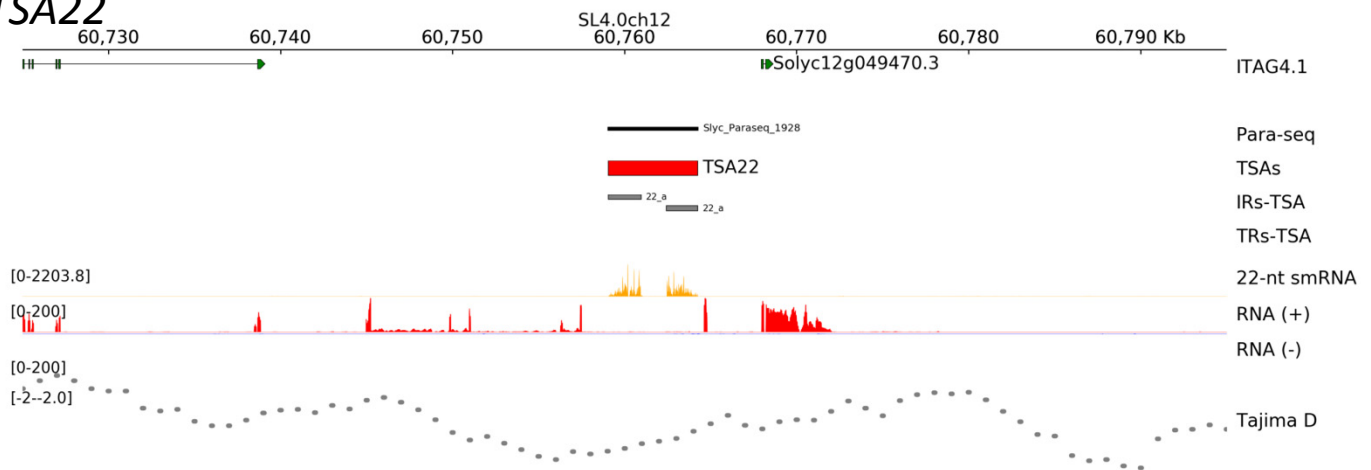


## TSA21



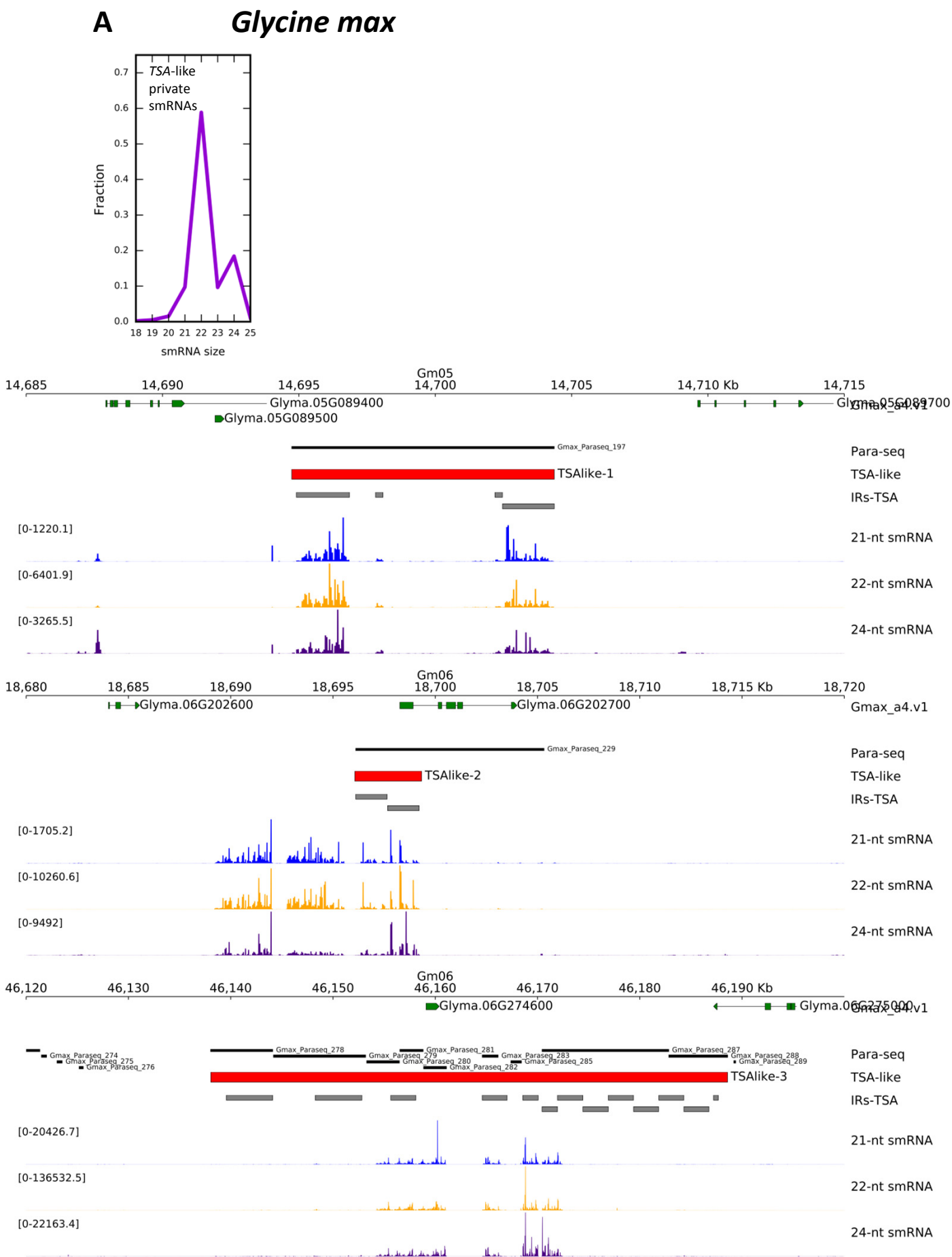
# Supplementary Figure S16

TSA22

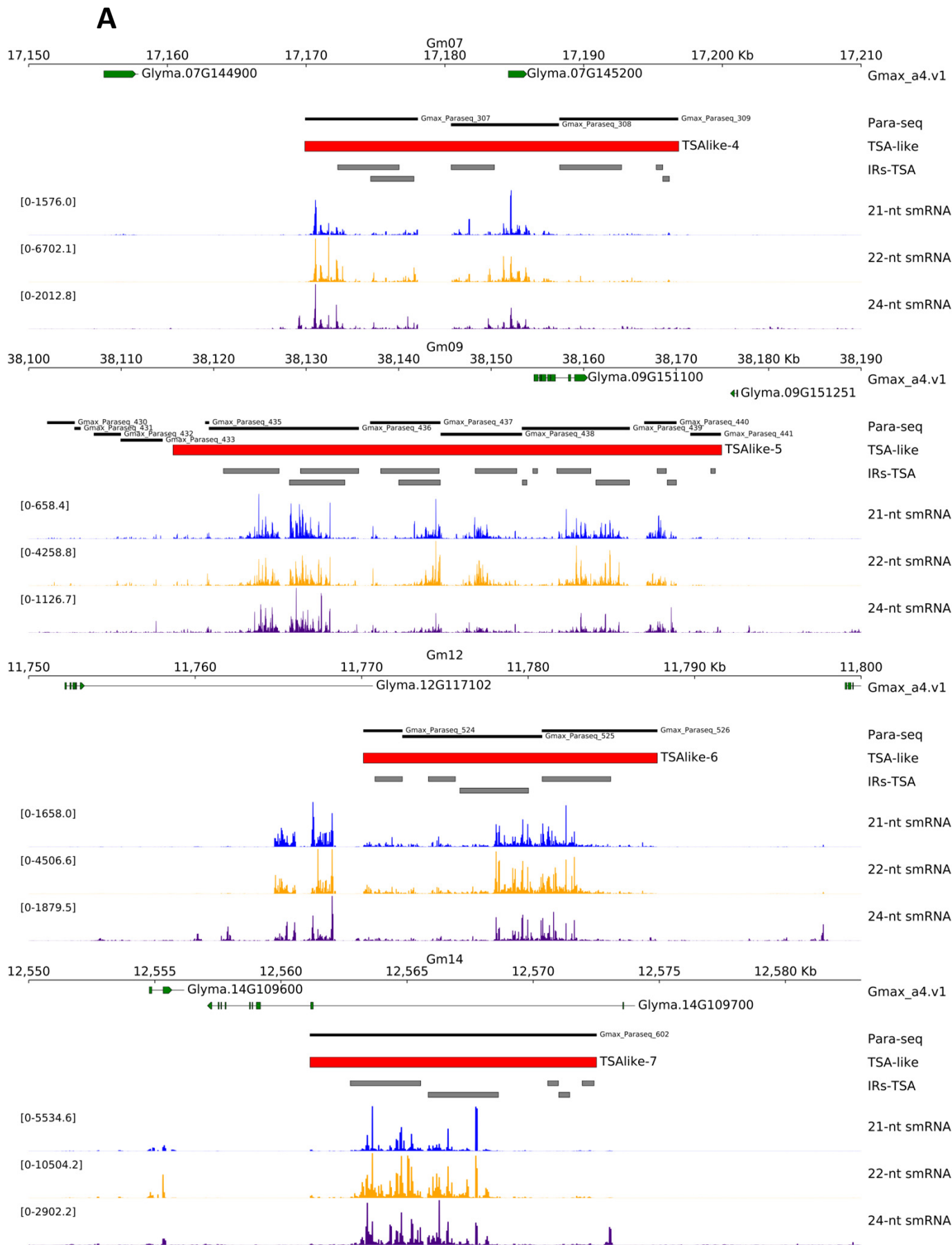


**Supplemental Figure S16:** Estimates of Tajima's D neutral summary statistic around TSAs coordinates inferred from private DNA-seq reads across 124 *S. lycopersicum* accessions. Window centres for 10000 bp sequences in 1000 bp sliding windows are shown (grey). On top, additional visualized data include ITAG4.1 gene models (green), recognized pararetroviral sequences (Para-seq, black), proposed positions for TSAs (red), inverted- and tandem-repeats in TSAs (IRs-TSA and TRs-TSA, grey and white), mapping siRNA signals for 22-nt (orange), and positive (red) and negative (blue) strands RNA-seq signals from pooled tissues (leaves, flowers and meristems).

Supplementary Figure S17

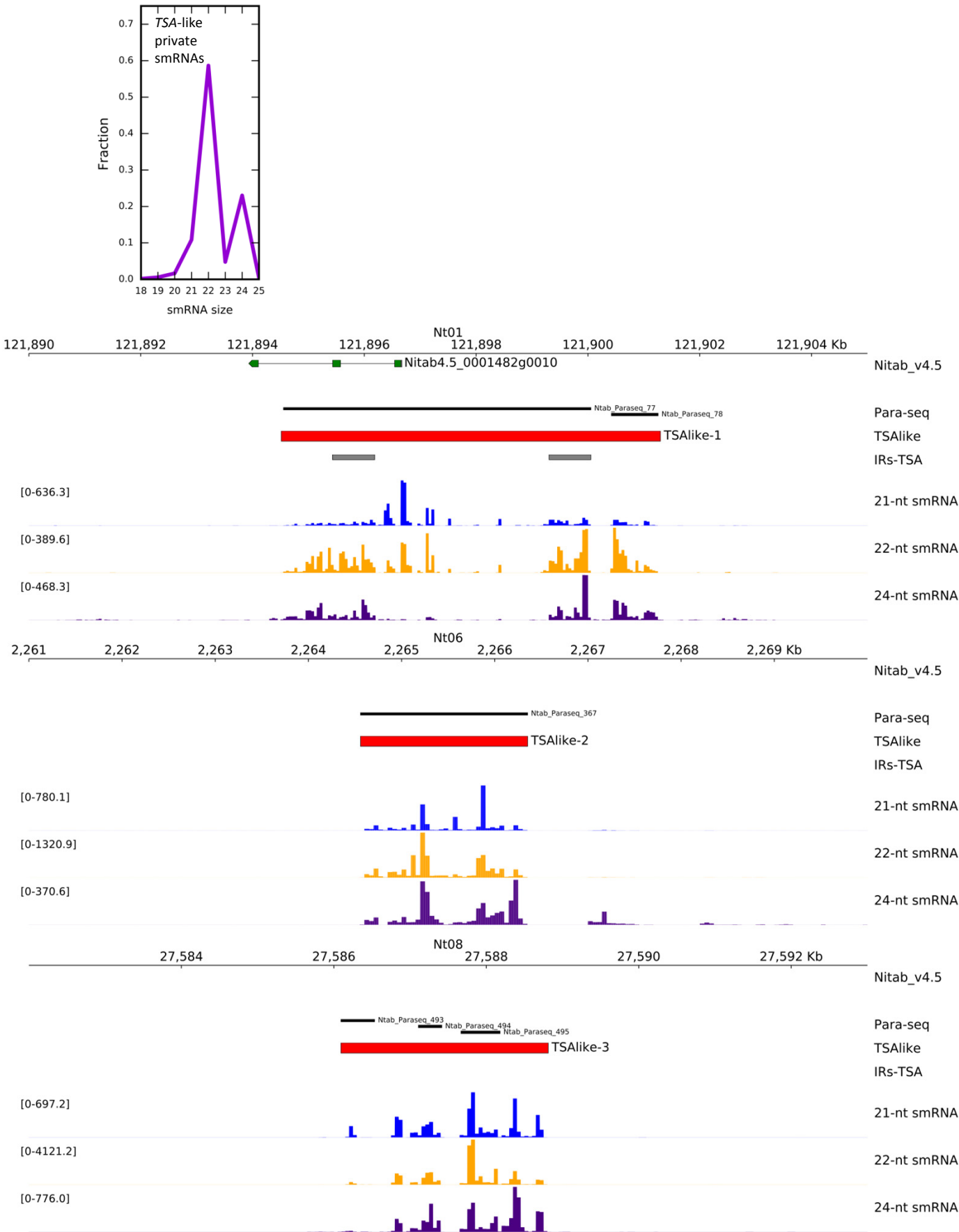


Supplementary Figure S17

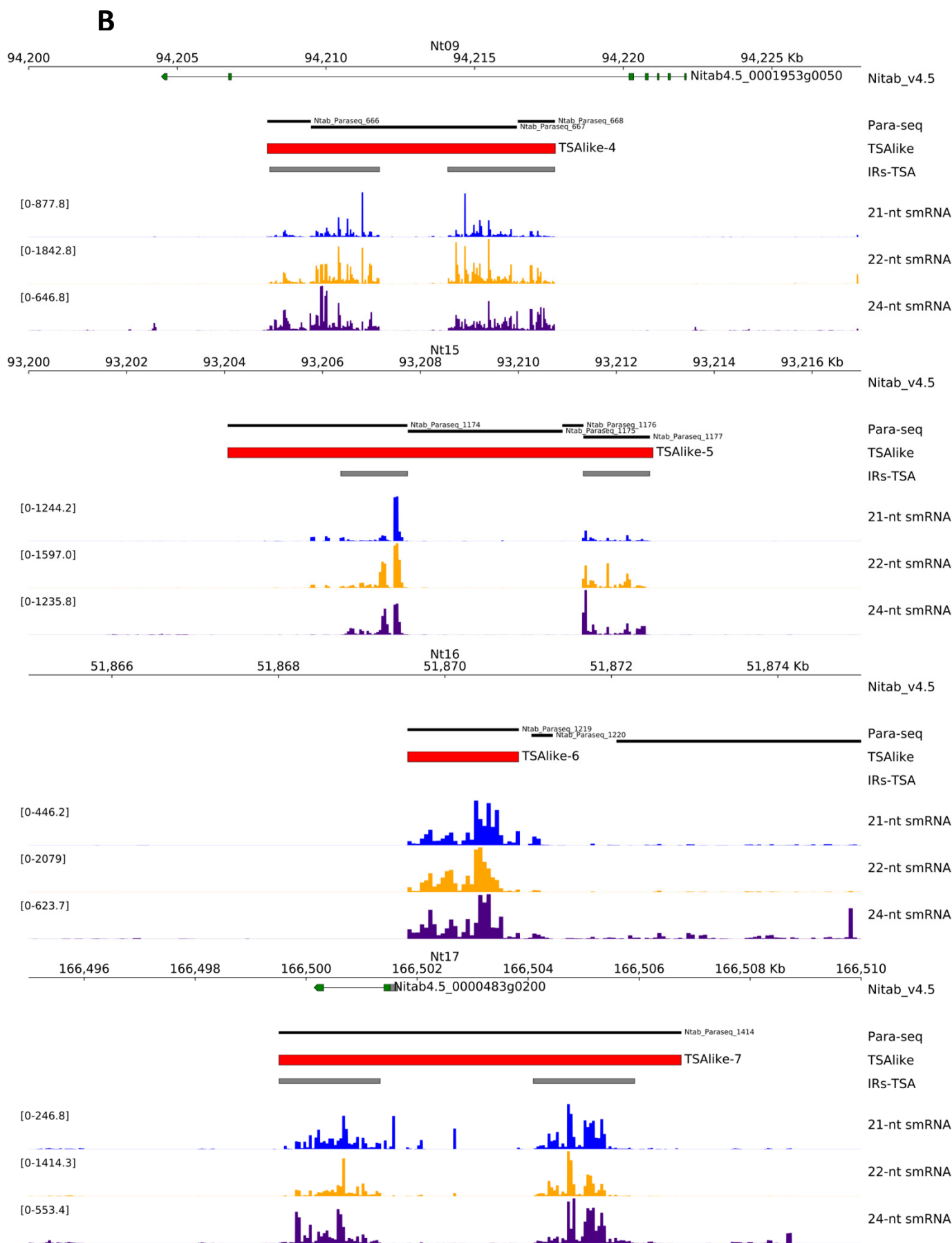


Supplementary Figure S17

**B** *Nicotiana tabacum*



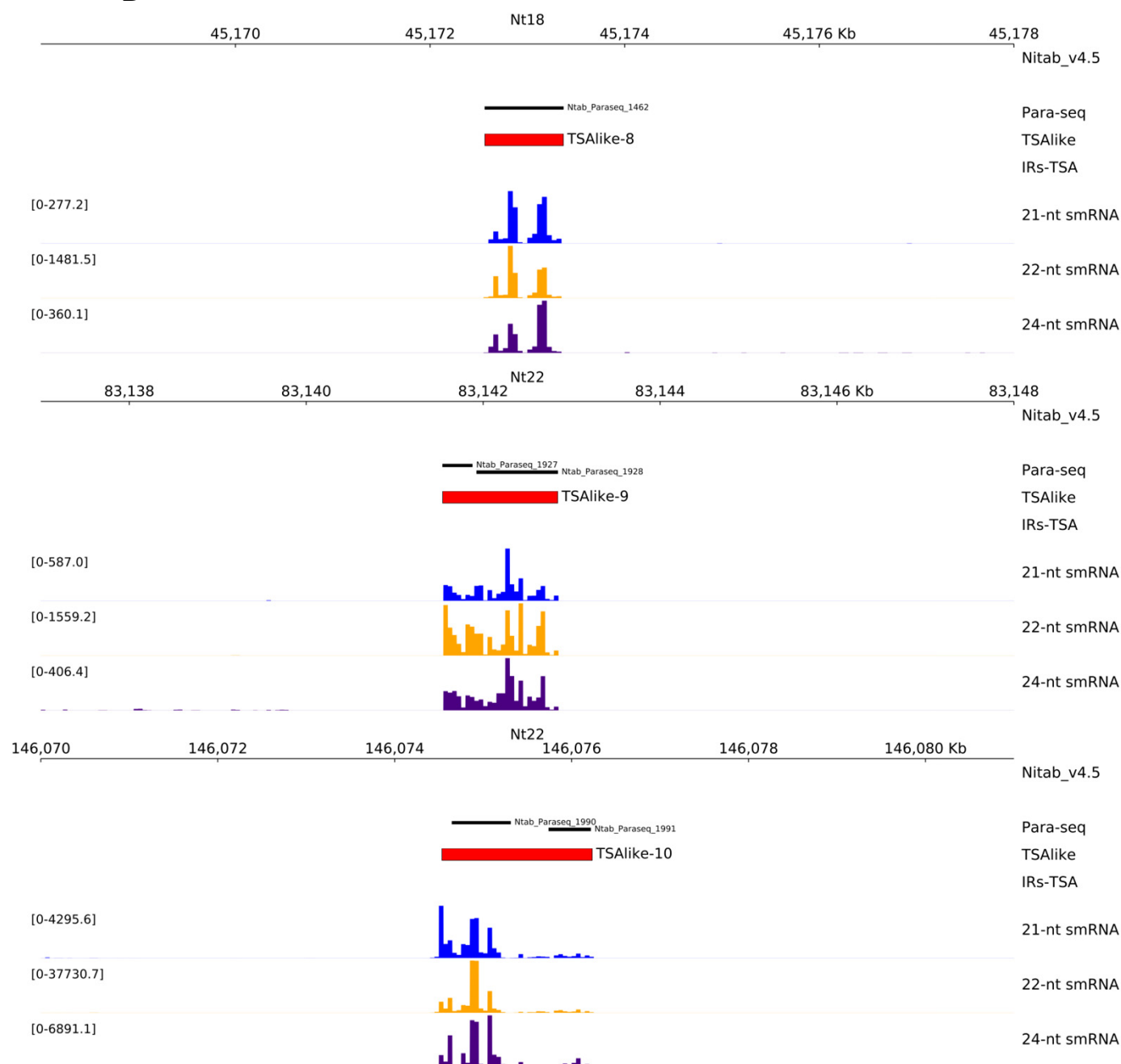
Supplementary Figure S17





# Supplementary Figure S17

**B**



**Supplemental Figure S17:** (A) Pooled private 18-25-nt siRNA library sizes mapping to currently recognized *G. max* TSA-like features (top), and genome-browser around these areas (bottom). (B) Pooled private 18-25-nt siRNA library sizes mapping to currently recognized *N. tabacum* TSA-like features (top), and genome-browser around these areas (bottom). Genome-browser data includes corresponding gene models (green), recognized pararetroviral sequences (Para-seq, black), proposed positions for TSA-like (red), inverted-repeats (IRs-TSA, grey) and mapping siRNA signals (21-nt = blue, 22-nt = orange and 24-nt = indigo).