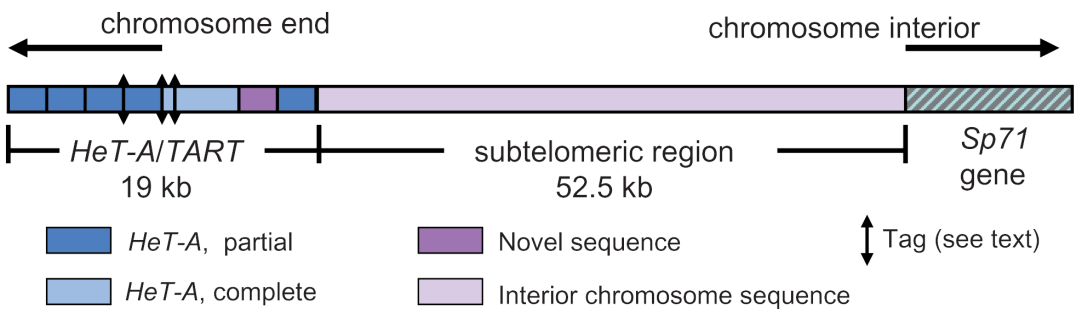
Supplemental Material

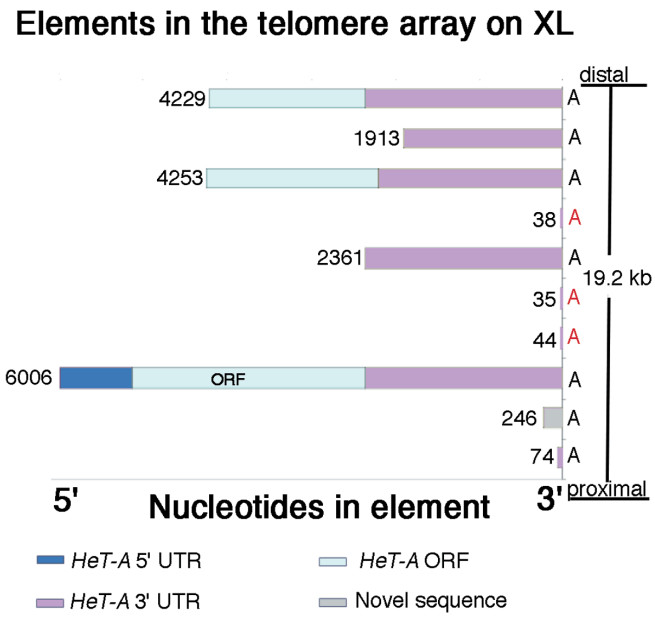
1. Euchromatic sites with similarity to *TART*

Sequence in the 3' UTR of *TART* A has similarity to sequence within the coding region of CG4118 (*nxf-2*) on chromosome 3L (nt 16542253→16542952). This similarity is in four segments, 59-227 nt in length, which are partially duplicated and rearranged. *TART* B has similarity to a cluster of 5 fragments in the *nej* coding region on the X chromosome (nt 9523390-9524208) but here also fragments are small (34-159 nt), scrambled and overlapping with respect to their order in *TART*. There are two other very small regions with similarity to *TART* B, 41 nt in chromosome 2R in CG13188 (nts 7481756→781797) and 25 nt in chromosome 3R (nts 18593299→18593284). Finally, there is one small region of similarity to the 3' UTR of *TART* C, 41 nt in the *Mnf* gene in 3L (11085460→11089501).

2. Sequence Organization on 4R and XL This section consists of Supplemental Figure 1 and Supplemental Tables 1 – 5.



Supplemental figure 1a



Supplemental figure 1b

Supplemental Figure 1. Elements in the *HeT-A* array in the telomere of XL in 2057. This depicts all of the currently assembled elements on XL (the most proximal 19.2 kb of this array). **(a)** Overview of elements in assembled sequence. Not drawn to scale. End of chromosome is on left. Right end shows the 52.5 subtelomeric sequence extending from the most proximal *HeT-A* element to the most distal gene, *Sp71.* This region contains many non-telomeric transposable elements (not shown here) All *HeT-A* elements are in the same orientation and all truncated elements have lost sequence from the 5' end. For details on size and sequence content of particular elements, see part **b**. **(b)** Each bar represents an element in the array drawn to scale with the total number of nucleotides indicated at the 5' end. Elements are aligned from the 3' end (at the right hand axis). In the array the 3'-most sequence of each element is actually connected to the 5' end of the element beneath it; but elements are shown separately for clarity. The most distal element in the assembled array is on top and the most proximal telomeric element is at the bottom. Open coding regions are light blue; complete coding regions are marked **ORF**. 3' ends marked **A** have an intact 3' end and oligo(A). Red **A** indicates elements thought to be "Tags" (see text).

Supplemental Table 1. Summary of Telomeric Elements on 4R and XL in 2057

Chromosome End & Element	Breakdown by class [a]	Totals	Com-plete	Partial	Tags	Truncated by cloning	Transition Region [b]
4R *HeT-A*	# of elements [c]	23	5	10	8		5
	% all *HeT* & Tart [d]	58.8	40.3	18.2	0.3	none	1.3
	% all *HeT-A* [e]	100	68.6	30.9	0.5		2.2
	% *HeT-A* coding [f]	100	87.0	13.0	none		none
4R TART	# of elements	8	2	5		1	2
	% all *HeT* & *TART*	41.2	34.1	4.0		3.0	1.1
	% all *TART*	100	82.9	9.8	N.A.	7.4	2.6
	% *TART* coding	100	100	none		none	none
XL *HeT-A*	# of elements	9	1	4	3	1	1
	% all *HeT* & *TART*	100	31.7	45.4	0.6	22.3	0.4
	% all *HeT-A*	100	31.7	45.4	0.6	22.3	0.4
	% *HeT-A* coding	100	41.6	30.4	none	27.9	none
XL *TART*	none	none	none	none	none	none	none

[a] The Totals column is a complete inventory of the telomeric retrotransposons *HeT-A* and *TART* in the most current assembly of chromosome ends in stock 2057. The Complete, Partial, Tags,

and Truncated-by-cloning columns show a breakdown by element class as discussed in the text. The Partial column excludes the most distal element if it is truncated by cloning.

[b] The transition region is defined here as the most proximal region of the *HeT-A*/*TART* array where the array has been invaded by other sequence. The Transition Region column is shown separately only to show its small size. Elements therein are also counted as Partials or Tags.

[c] Including most distal element in each array which is assumed to be truncated by cloning.

[d] Percentage of the entire sequence of *HeT-A* and *TART* in the specified class.

[e] Percentage of all *HeT-A* (or *TART*), including most distal element, found in the specified class.

[f] Percentage of total *HeT-A* (or *TART*) coding sequence found in the specified class.

Supplemental Table 2. Percent nucleotide identity [a] of complete *HeT-A* elements

Element [b]	23Zn-1	4-5892	4-5848	4-6012	4-5840	4-5570
4-5892	72.4					
4-5848 [c]	71.9	76.4				
4-6012	73.2	90.9	78.1			
4-5840	72.1	76.7	99.2	77.9		
4-5570	82.7	69.8	67.11	68.5	67.5	
X-6006	73.4	76.5	79.1	78.1	79.0	68.2

[a] Percent identities were determined from global alignments made by the LALIGN program (www.ch.embnet.org/software/LALIGN_form.html).

[b] Element names contain chromosome location, 4 or X plus the total number of nts in that element in 2057, except for 23Zn-1, which is the 6083 nt canonical sequence for *HeT-A*.

[c] This element lacks the terminal GTT and may not be competent to transpose.

Supplemental Table 3. Percent amino acid and nucleotide identity [a] of *HeT-A* coding regions. (Nucleotide % in parentheses and bolded.)

Element [b]	23Zn-1	4-5892	4-5848	4-6012	4-5840	4-5570
4-5892 *(944)* [c]	78.6 [d] **(81.7)**					
4-5848 *(951)* [e]	77.5 **(81.9)**	81.5 **(84.8)**				
4-6012 *(944)*	77.9 **(81.5)**	93.8 **(95.4)**	80.1 **(84.1)**			
4-5840 *(951)*	77.4 **(81.8)**	81.5 **(84.8)**	99.7 **(99.7)**	80.1 **(84.7)**		
4-5570 *(921)*	96.2 **(96.6)**	77.6 **(81.6)**	77.7 **(82.7)**	78.5 **(82.1)**	77.6 **(82.7)**	
X-6006 *(929)*	76.1 **(79.7)**	76.6 **(81.5)**	81.9 **(83.9)**	75.5 **(81.1)**	81.8 **(83.9)**	75.9 **(79.9)**

[a] Percent identities were determined from global alignments made by the LALIGN program (www.ch.embnet.org/software/LALIGN_form.html).

[b] Element names contain chromosome location, 4 or X, plus the total number of nt in that element in 2057, except for 23Zn-1, which is the canonical sequence for *HeT-A* and encodes 921 amino acids.

[c] Numbers in italics are the number of amino acids in each coding region.

[d] Upper number is percent of amino acid identities. Lower number is percent of nucleotide identities (in parentheses and bolded). Note that, with one exception, there are fewer amino acid than nucleotide identities.

[e] This element lacks the terminal GTT and may not be competent to transpose.

Supplemental Table 4. Per cent nucleotide identity [a] between elements in *TART* subfamilies

TART Subfamily	A (Canton S)	B	C
A (Oregon R)	99.3	61.7	70.0
B			68.7

[a] Percent identities were determined from global alignments made by the LALIGN program (http://www.ch.embnet.org/software/LALIGN_form.html)

[b] Elements used were: A (Oregon R), AY561850: A (Canton S), AJ566116; B, U14101; C, AY600955. The 5' ends of these elements have not been completely defined but, because the 5'UTR is identical to the PNTR sequence in the 3' UTR of the same element, we have omitted 5' UTR sequence in all comparisons.

Supplemental Table 5. Per cent amino acid and nucleotide identity [a] in coding region of *TART* subfamilies.

(Nucleotide % in parentheses and bolded.)

TART Subfamily	A (Canton S)	B	C
A (Oregon R)	ORF1=99.9 **(100)**	ORF1=82.9 **(85.2)**	ORF1=82.5 **(85.4)**
	ORF2=99.9 **(99.9)**	ORF2=91.3 **(88.0)**	ORF2=90.9 **(89.1)**
B			ORF1=94.7 **(95.5)**
			ORF2=89.9 **(88.4)**

[a] Per cent identities were determined as described for Supplemental Table 4, using the elements listed there.

3. Are the number of *HeT-A* and *TART* elements related across lineages and tissue type?

Correlation in the number of retroelements. The most obvious pattern displayed in Figure 4 is the surprisingly apparent codependence of *HeT-A* and *TART* over all lineages examined. To investigate this phenomenon, and to test its reality, i.e., to check how frequently the obvious relationship could result from random error in our measurements, we grouped each lineage together in various ways and analyzed the *HeT-A* to *TART* relationship for each group by correlation analysis. Group 1 included all head data; group 2, all salivary gland data; group 3, all head and all salivary gland data; Group 4: all head, all salivary gland, and S2 cell data. Supplemental Table 6 shows that there is significant linear correlation of the number of *HeT-A* elements with the number of *TART* elements in all groups examined, and that the likelihood of this happening by chance is very small. Specifically, we conclude that the numbers of *HeT-A* and *TART* in

any lineage are significantly correlated in each grouping at confidence levels (C.L.) ranging from 93% to > 99%. Confidence levels increase as the number of samples in the group increases because more samples reduce statistical uncertainties.

Supplemental Table 6. Linear Correlation of HeT-A with TART over all lineages

	r	*t*	*p*	C.L.
HeT-A with TART (heads only)	0.96	6.66	0.02	98%
HeT-A with TART (salivaries only)	0.88	3.65	0.07	93%
HeT-A with TART (both heads and salivaries)	0.94	7.63	0.0003	>99%
HeT-A with TART (heads, salivaries, S2 cells)	0.94	8.25	0.0001	>99%

r is the linear correlation coefficient for the data sets in each row; *p* is the probability that the data sets are *not* correlated, i.e. the probability that the correlation coefficient could be so large by chance, evaluated by Students 2-sided t-test; *t* is the "statistic" used to evaluate correlation ($t = r\times[(N-2)/(1-r^2)]^{1/2}$) with *(N – 2)* degrees of freedom; *N* is the number of data points in each set being compared, and the confidence level, C.L.=*(1-p).*

Under-replication in salivary glands. Another pattern shown in figure 4 is that there seem to be fewer *HeT-A* and *TART* elements in larval salivary glands than in female adult heads and that there appears to be a near-linear relationship across stocks between the number of *HeT-A* elements in heads and the number of *HeT-A* elements in salivary glands; *TART* numbers also appear to track each other, although less obviously.

To investigate these phenomena we grouped the data, this time by tissue and retroelement type. Supplementary Table 7 shows the results for correlation of *HeT-A* and *TART* in heads vs salivary glands with confidence levels similar to the *HeT-A* vs. *TART* correlation. Once again the confidence level increased with sample size.

Supplemental Table 7. Linear Correlations of HeT-A and TART over all stocks.

	r	*t*	*p*	C. L.
HeT-A heads with *HeT-A* salivaries	0.99	9.83	0.01	99%
TART heads with *TART* salivaries	0.95	4.11	0.05	95%
HeT-A & *TART* heads with *HeT-A* &*TART* salivaries	0.97	10.50	0.00004	>99%

Replication Ratios. We define the replication ratio to be the quotient of the number of complete elements per genome of each stock in the salivary glands divided by the number of the same element in the adult head. Values of the replication ratios are given in the main text, Fig. 5 and Table 1. As can be expected from the display in Fig. 5, *HeT-A* and *TART* replication ratios (*HeT-A* vs. *TART* replication ratios across stocks) are significantly correlated across stocks: r=0.67, p = 0.009, C.L. = 99%

However, it is more important to know if the *HeT-A* and *TART* replication ratios are likely to have the same average value; they do not. To evaluate the possibility that they have the same averages, we compared the *HeT-A* and *TART* ratios for all stocks

and enzymes using Student's heteroscedastic t-test for significantly different means and found the probability of the *HeT-A* and *TART* means being equal to be small, $p = 0.11$. A full ANOVA analysis gives the same result, the *HeT-A* and *TART* averages differ significantly at the 89% confidence level. Even a non-parametric (see Data Analysis, below) Kruskal-Wallis ANOVA analysis reported difference in the median values with 71% confidence level. Thus, the observed under-replication of *TART* is significantly less than the under-replication of *HeT-A.*

4. Data Analysis

Dose curves. Gels used to determine the absolute number of coding region equivalents per genome also contained three lanes each ("1X", "2.5X", and "5X") of the *rp49* and retrotransposon probes, from which we determined a dose curve for each probe. (A 1X probe lane was loaded with the same number of probe DNA molecules as the number of genomes loaded in each sample lane, assuming the length of the genome to be $1.8 \cdot 10^8$ bp.) For each experiment the probe lane data were fitted to give the best estimate of mean 1X probe count. Each lane of genomic DNA measurements was normalized by the *rp49* measurement of the same lane and multiplied by the ratio of the mean 1X *rp49* probe to the mean 1X genomic probe. The result is the number of ORF equivalents per genome of that stock, i.e., the total length of ORF sequence of that stock in the genome, both full length and partial, measured in units of ORF coding sequence. For graphical and tabulated display of results, these values were averaged over the restriction enzymes, and measurement uncertainty of these averages was estimated by the standard error of the entire sample being averaged. More complex analyses to investigate variation with genotype and/or tissue, used individual restriction enzyme data, not averages, expressed in an appropriate matrix by enzyme, by stock and/or tissue, and, for *HeT-A* to *TART* comparisons, by retrotransposon.

Statistical analysis of the data. In all cases, we accounted for the possibility of non-normal distribution of measurement errors by comparing parametric analyses, which assume that errors are normally distributed, with non-parametric analyses, which avoid the assumption of normality. The results were comparable and significance is claimed only if both approaches agreed (although confidence levels differed somewhat). Except as noted, differences in lineages were generally calculated using one-way Pearson analysis of variance (ANOVA) and non-parametric Kruskal-Wallis ANOVA; In these analyses, protection against assuming different stocks are significantly different when in

fact they are not was provided by application of Tukey's or Dunnet's algorithms, as appropriate, at the 95% confidence level and by Bonferroni's algorithm at lower confidence levels. Correlation calculations used both parametric Pearson and non-parametric Spearman rank correlation analysis. All non-parametric analyses were performed using the Analyse-it add-in package for Excel (Analyse-it for Microsoft Excel, Leeds, UK. See http://www.analyse-it.com/).