

Figure S1

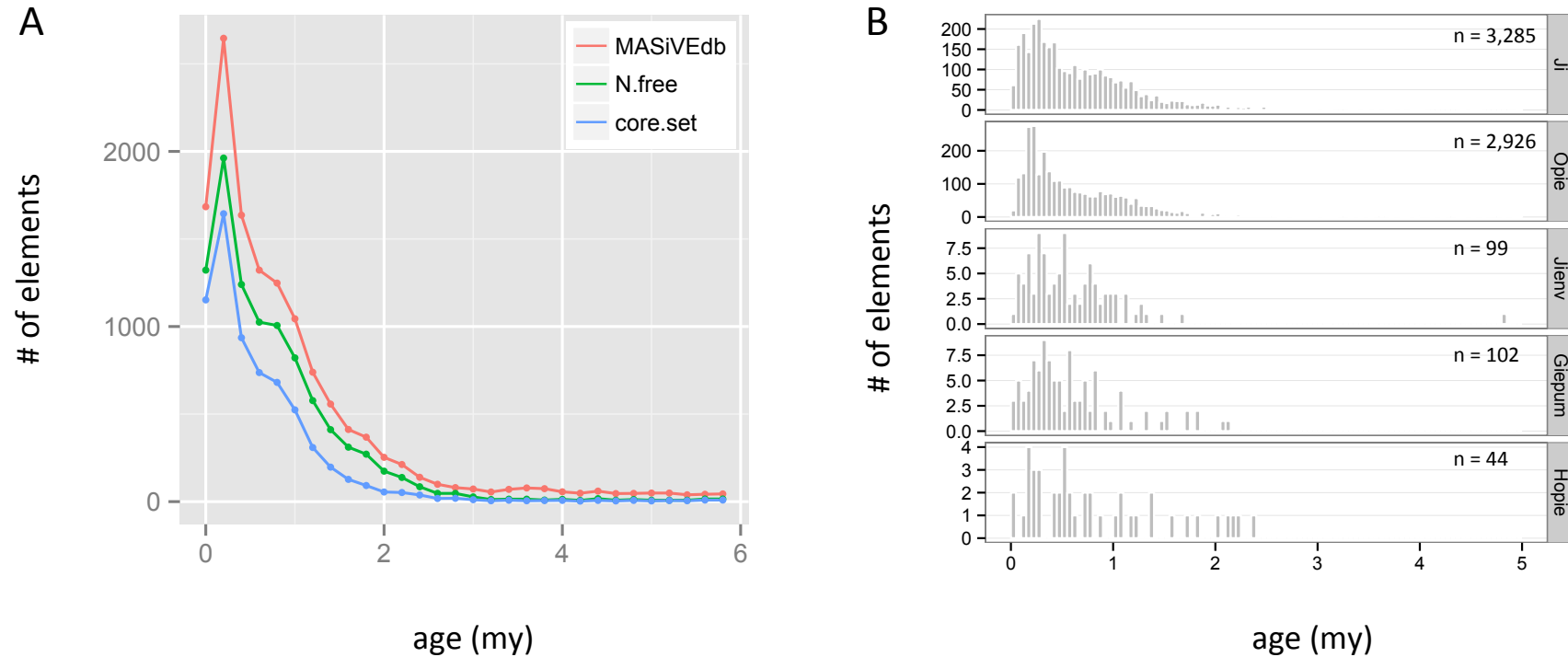


Figure S1. Sirevirus insertion age. **(A)** Age distribution in 0.2 million years (my) windows of all 13,833 elements of MASiVEdb, of 9,806 elements that passed the first filter of our pipeline ('N.free', not to contain >5 'N' nucleotides in their sequence, see Methods), and of the final set of 6,456 elements. The x-axis is cutoff at 6my. **(B)** Age distribution in 0.05my windows of the five families that comprised the final set. The x-axis is cutoff at 5my. Due to the high range of numbers in the y-axis and the age windows in the x-axis, elements older than ~2-2.5my cannot be identified in these figures. See Table S4 for the distribution of Sireviruses across age groups.

Figure S2

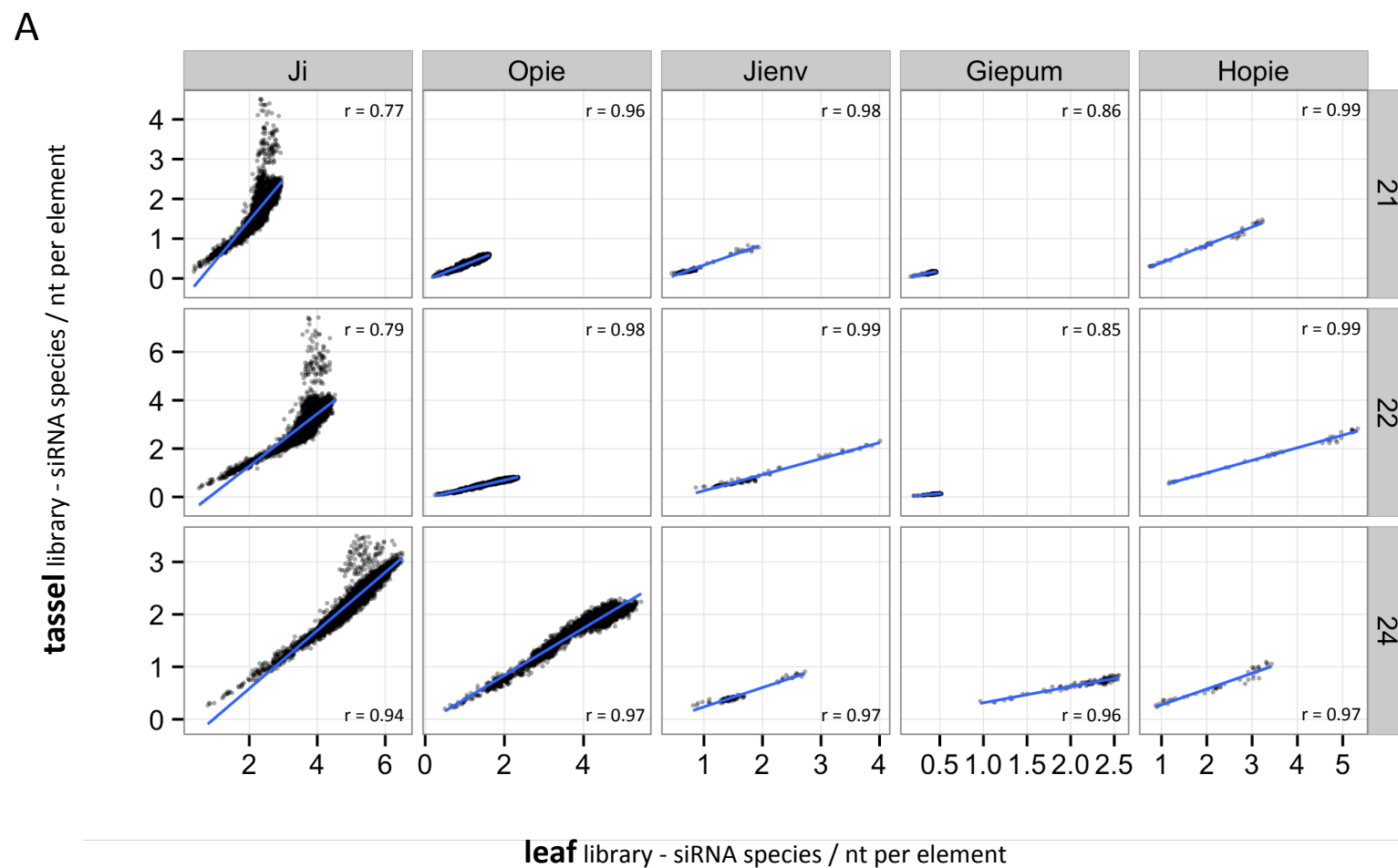


Figure S2. Cross-library comparison of siRNA targeting to individual elements. Targeting was calculated per nucleotide of the full-length sequence for each siRNA length and family separately. **(A)** leaf vs. tassel libraries; **(B)** leaf vs. immature ear libraries; **(C)** tassel vs. immature ear libraries. The grey shading around each regression line represents the 0.95 confidence interval level, while the Pearson r is shown for each plot (P values were $<10^{-20}$ in nearly all cases).

Figure S2 (cont.)

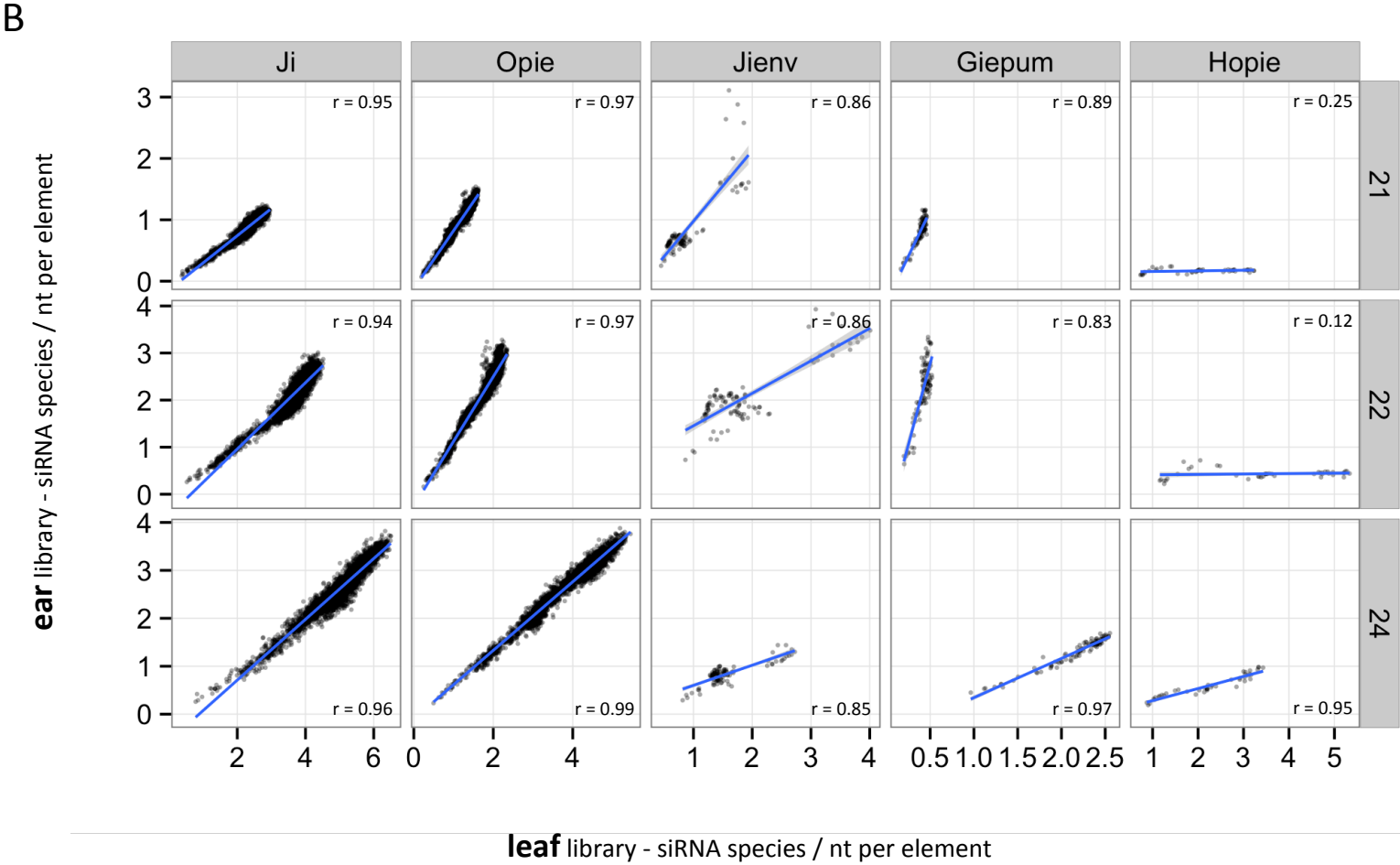


Figure S2 (cont.)

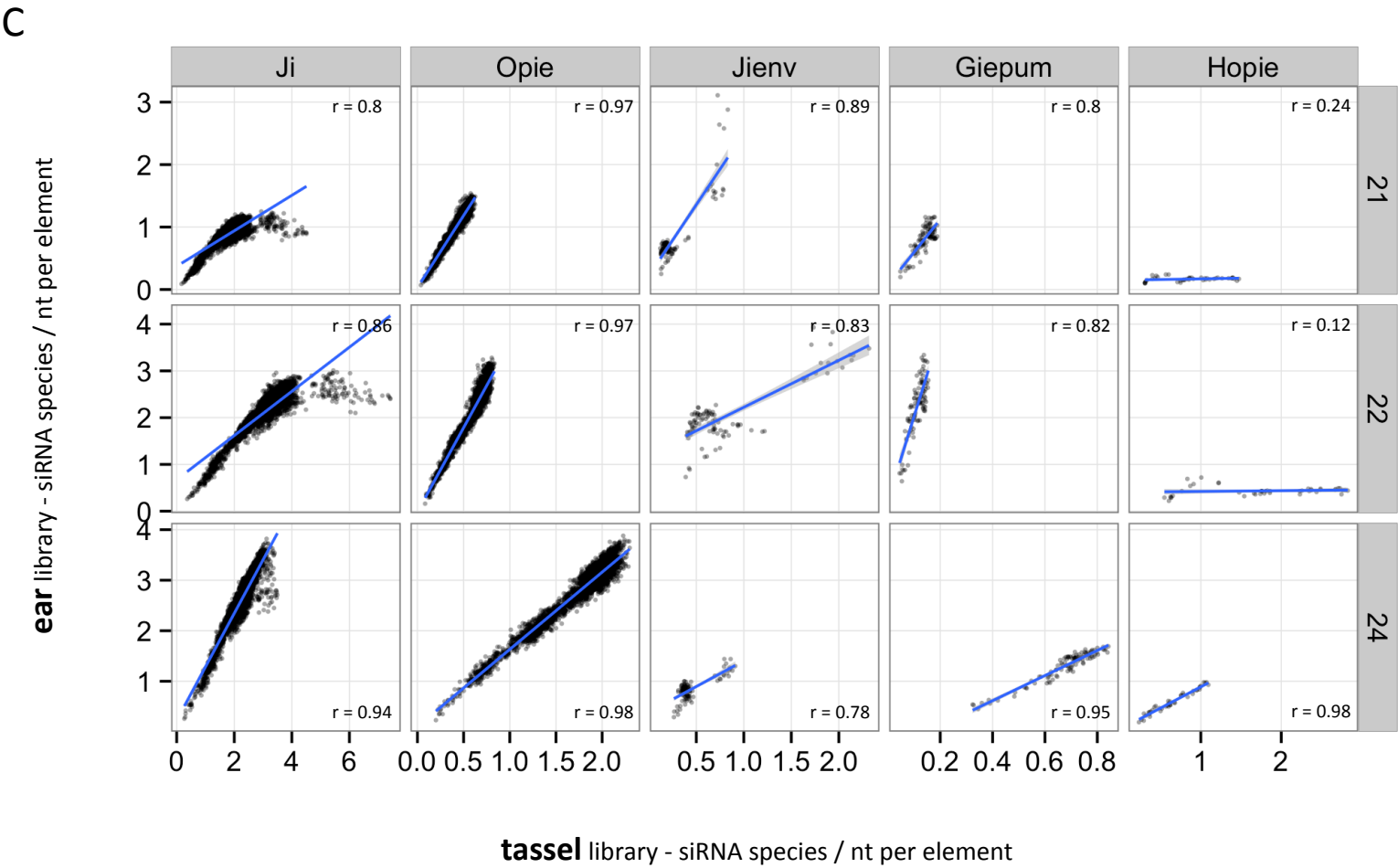


Figure S3

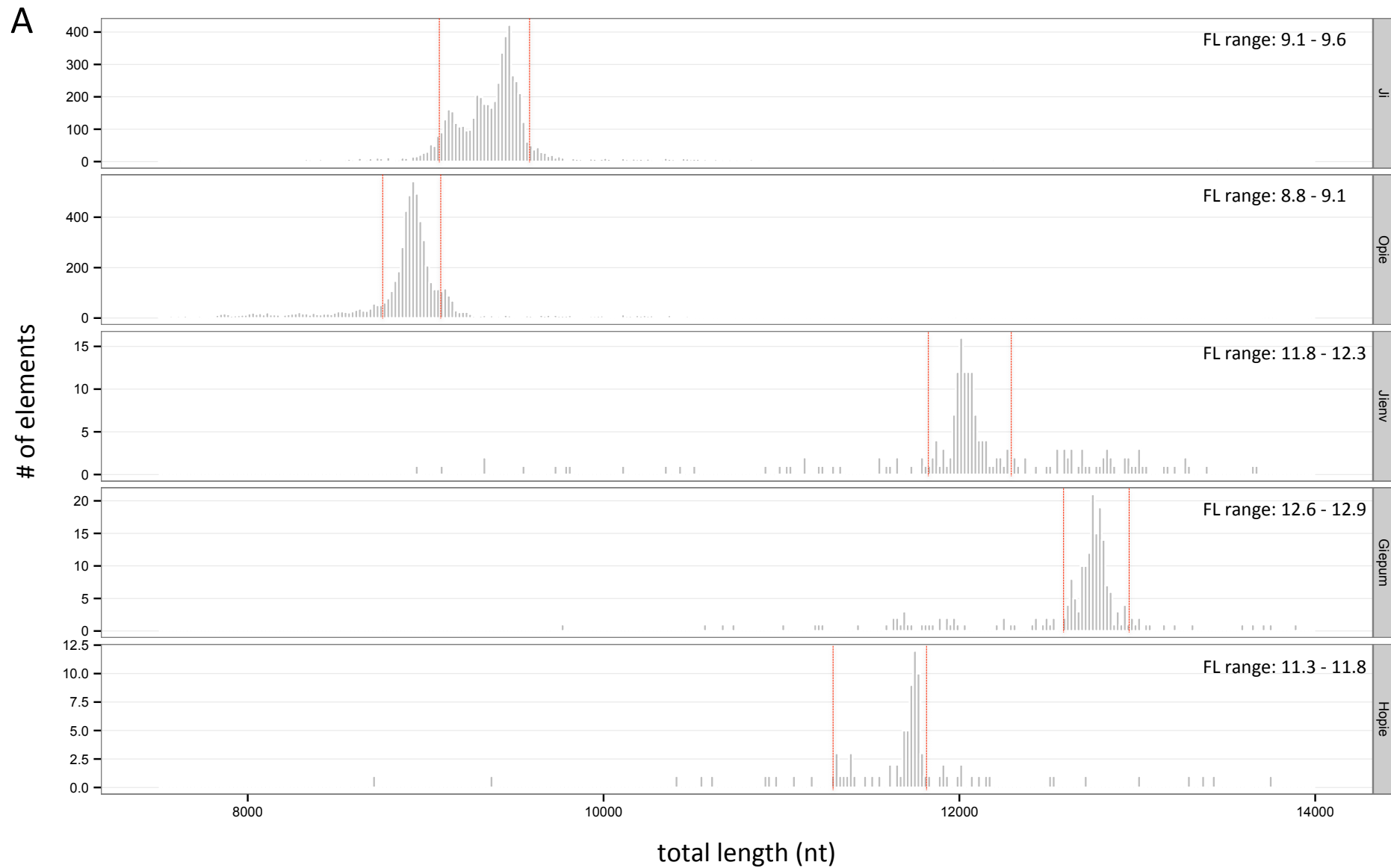


Figure S3. Sirevirus length analysis. **(A)** Total and **(B)** LTR lengths of all 13,833 maize elements that belong in five Sirevirus families. The dotted red lines approximately indicate the ranges (also shown in kb) used in this study for producing the set of 6,456 elements.

Figure S3 (cont.)

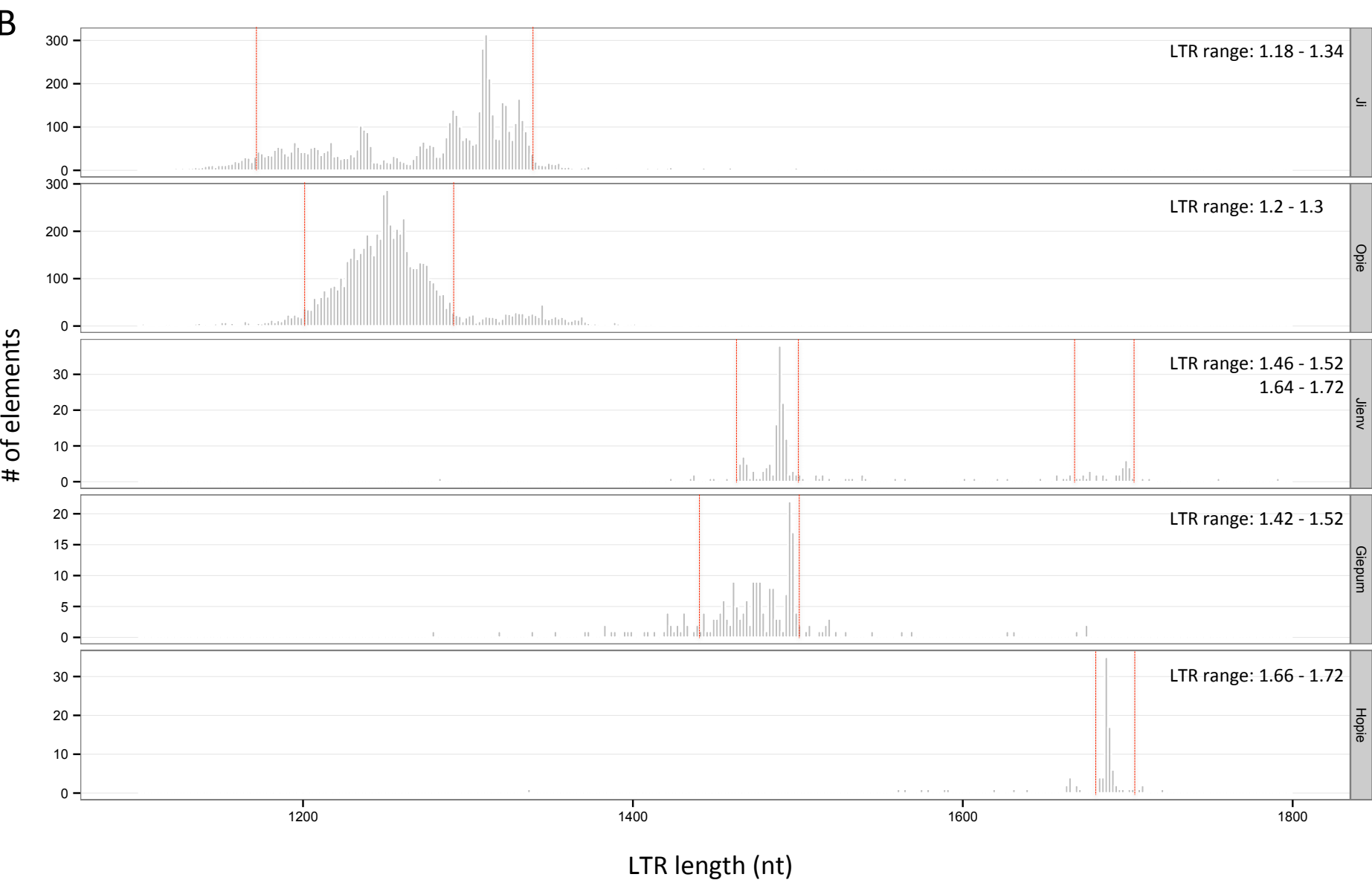


Figure S4

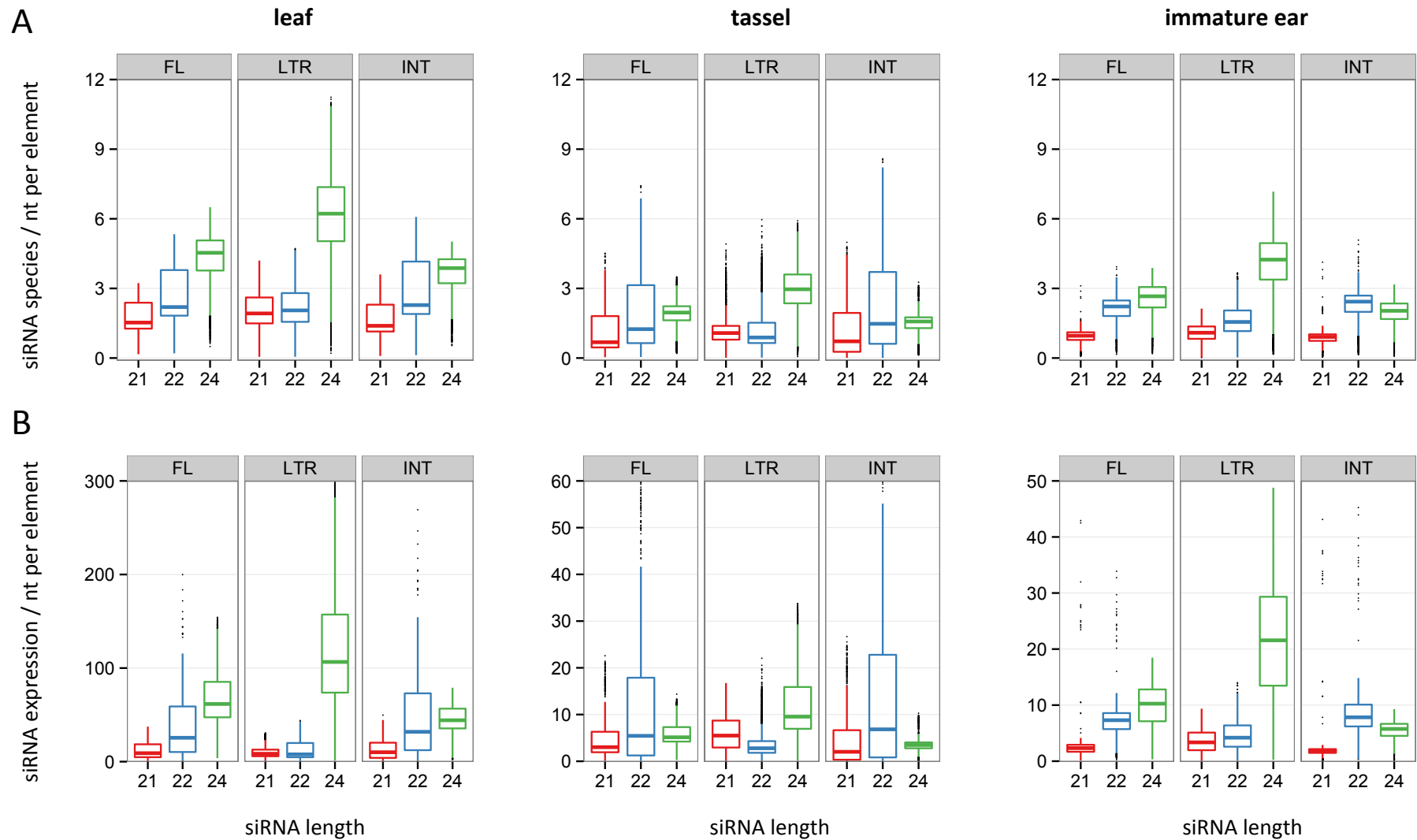


Figure S4. General siRNA targeting patterns of Sireviruses using data from all three libraries. Number of siRNA species (**A**) and expression (**B**) per nucleotide of the full-length (FL), LTRs or INT domain of all 6,456 Sireviruses.

Figure S5

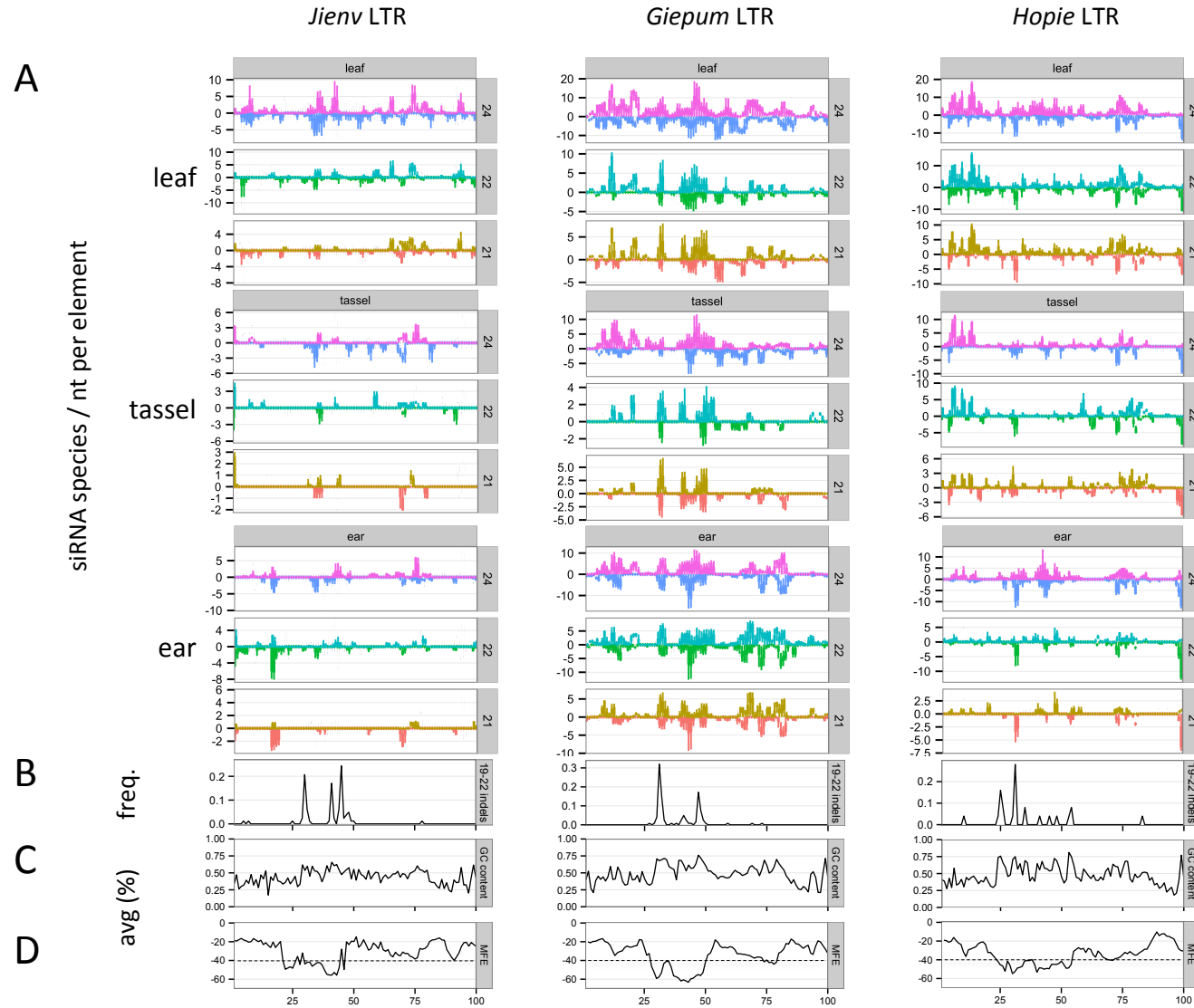


Figure S5. Epigenetic and sequence patterns of *Jienv*, *Giepum* and *Hopie* elements. Each LTR and INT domain was first split in 100 equally-sized windows of ~15nt and ~92nt length respectively. **(A)** Number of siRNA species per nucleotide along the sense (positive x-axis) and antisense (negative x-axis) strands of the LTRs and INT domain of all elements for each family, visualized with a boxplot for each window. **(B)** Distribution of 19-22nt long indels. **(C)** Average GC content. **(D)** Average minimum free energy (MFE) of the siRNA loci (20nt upstream + siRNA length + 70nt downstream as in Wang et al. (2009)) across the genomes of the three families. The miRNA cutoff of -40 MFE is highlighted by dotted lines. For the INT domain, the approximate positions of the *gag*, *pol* and *envelope*-like genes are shown.

Figure S5 (cont.)

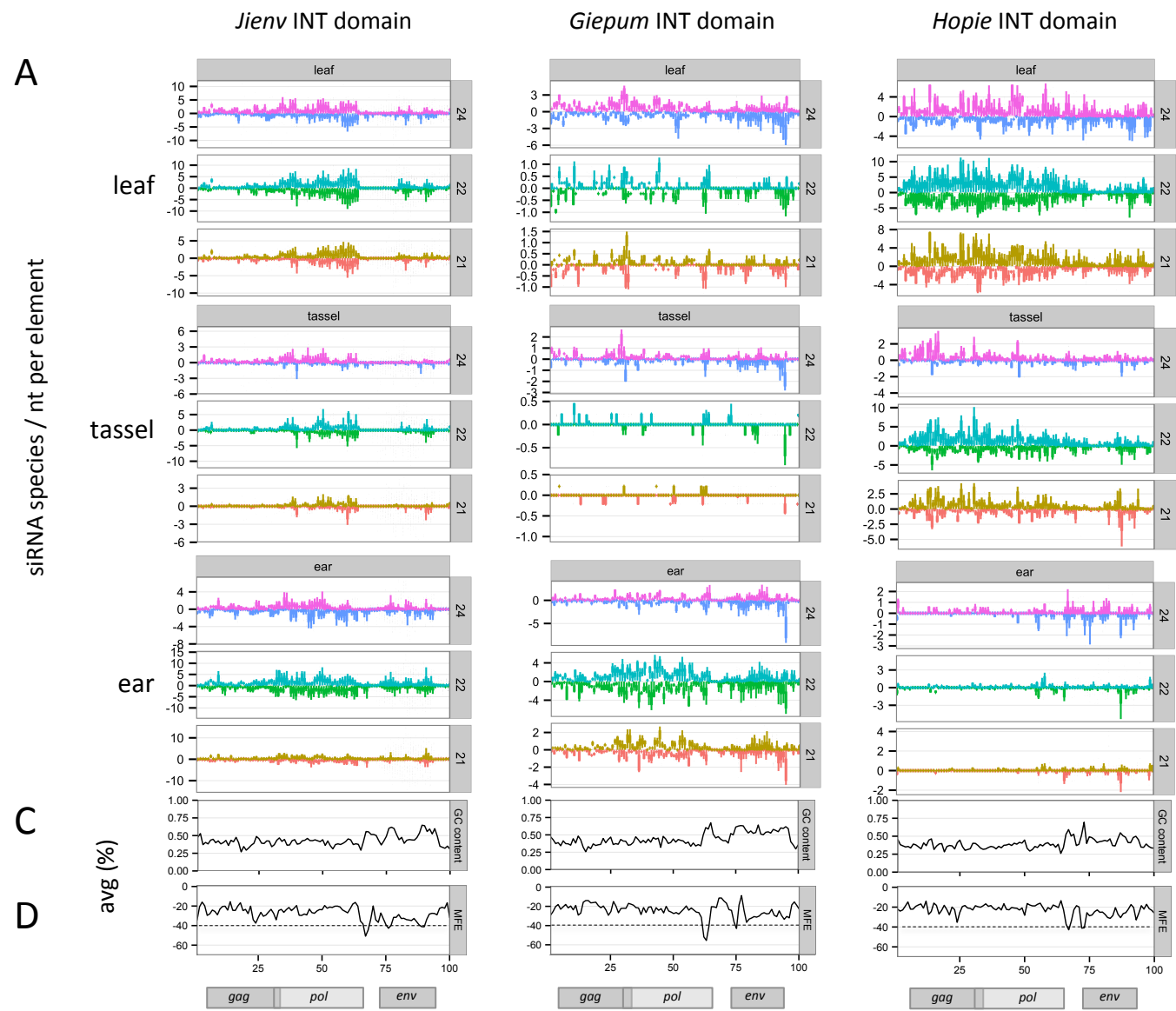


Figure S6

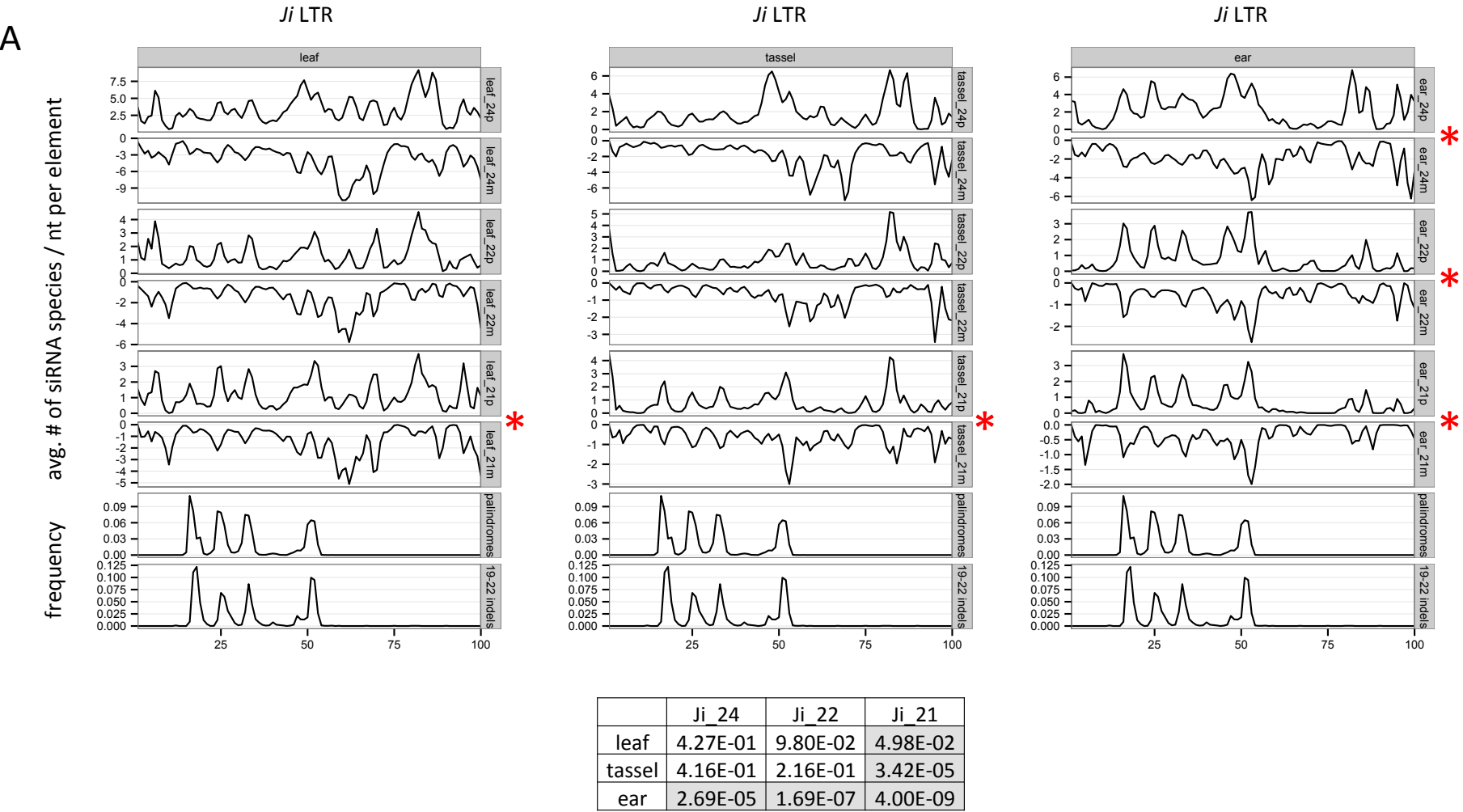
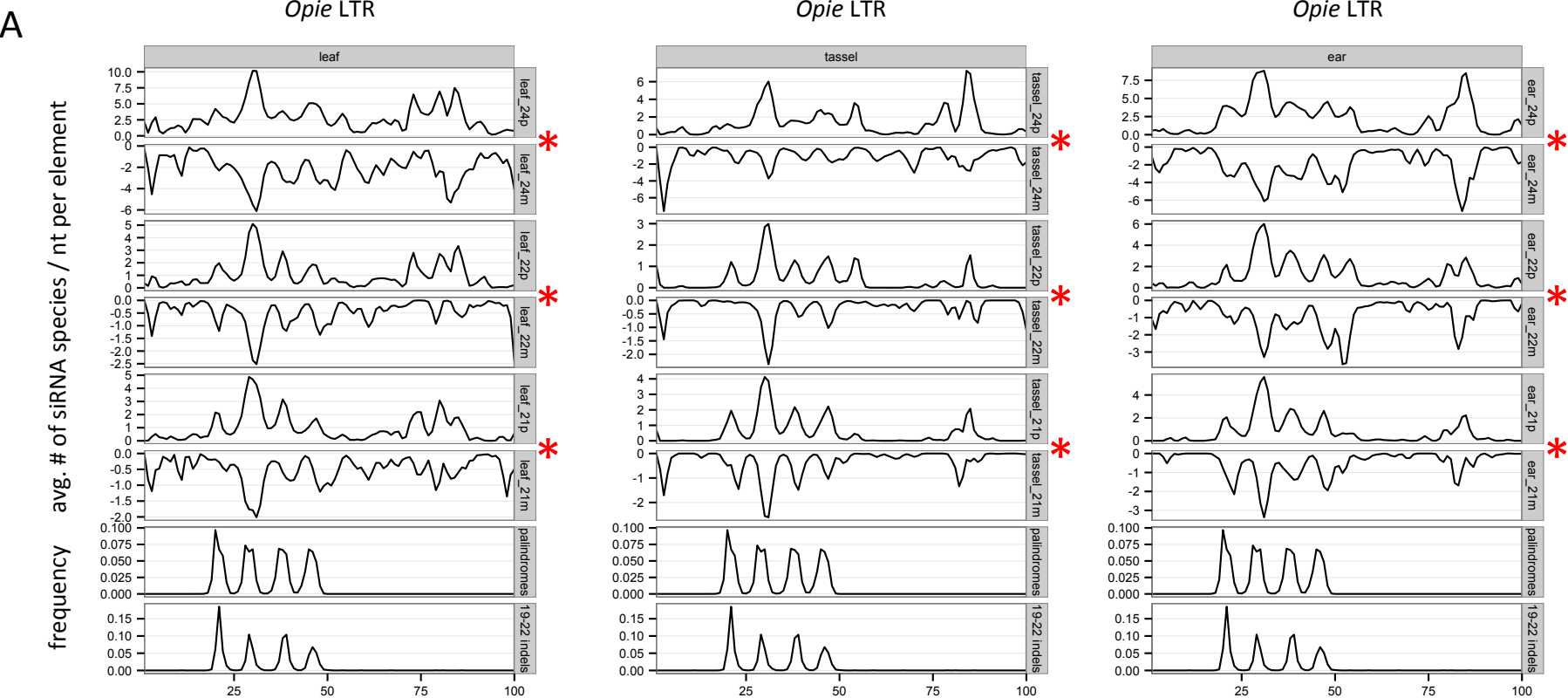


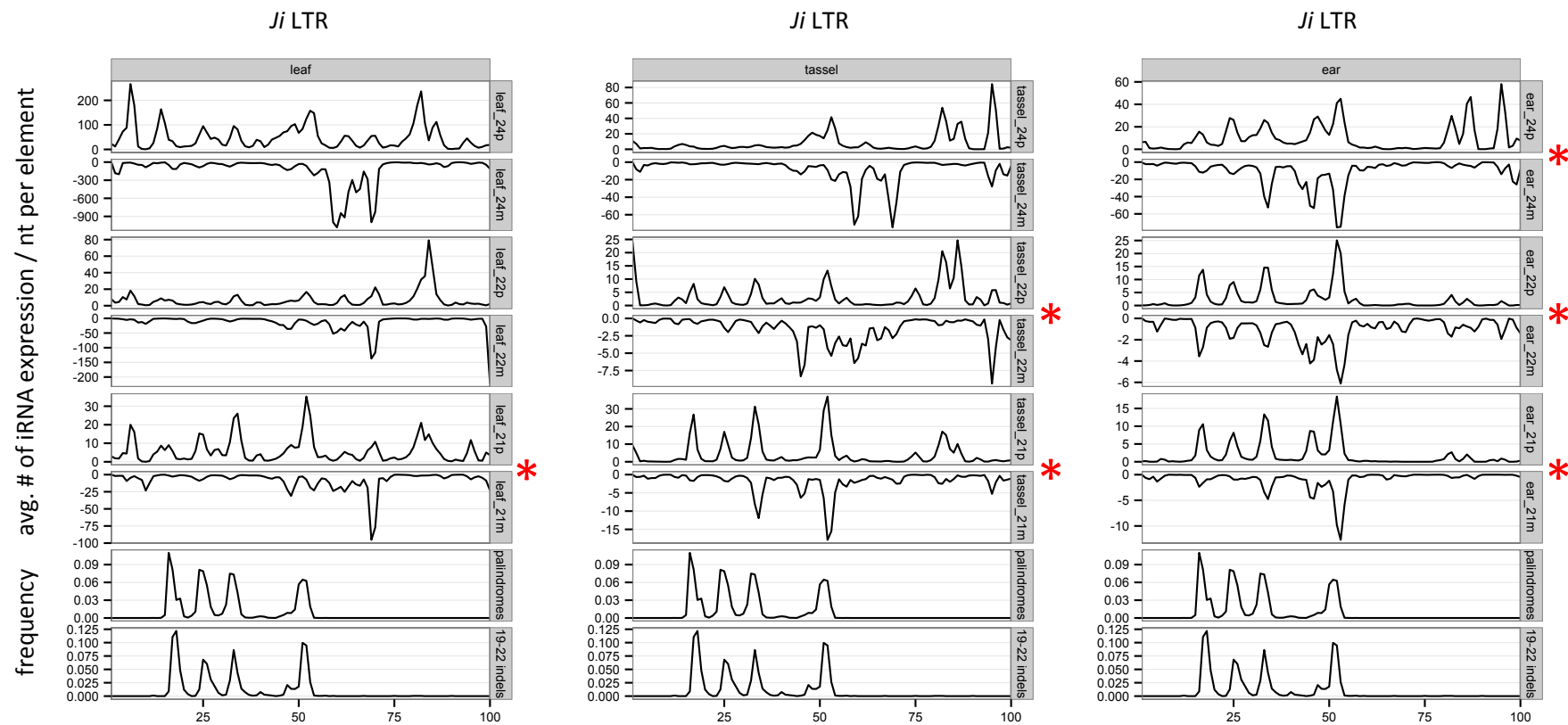
Figure S6. Statistical analysis of the siRNA/palindrome loci overlap using different types of siRNA data. **(A)** Number of siRNA species per nucleotide along the sense (positive x-axis) and antisense (negative x-axis) strands of the LTRs of 3,285 *Ji* and 2,926 *Opie* elements. Each LTR was first split in 100 equally-sized windows of ~13nt. The plots are identical to those of Figure 2A, but here the average of each window is depicted instead of a boxplot that contains all measurements. Red asterisks indicate plots where the overlap was statistically significant according to the Mann-Whitney U test (*P* values adjusted by Bonferroni correction). For the test, the four palindrome loci were combined, hence representing 16% of the LTR length, and tested for a significantly higher number of siRNA species per nucleotide and for both strands combined compared to the rest of the LTR. The bottom two plots of each panel show the distribution of palindromes and their 19-22nt indels. The table contains the *P* values of the Mann-Whitney U test for each combination of family, siRNA length and tissue. Cells highlighted in grey depict the statistically significant *P* values. **(B)** As in (A) but using siRNA expression measurements. **(C, D)** As in (A, B) but using siRNA data that were first normalized by their number of genomic locations. Applying all above tests on the sense and antisense strands separately, produced highly similar results (not shown).

Figure S6 (cont.)



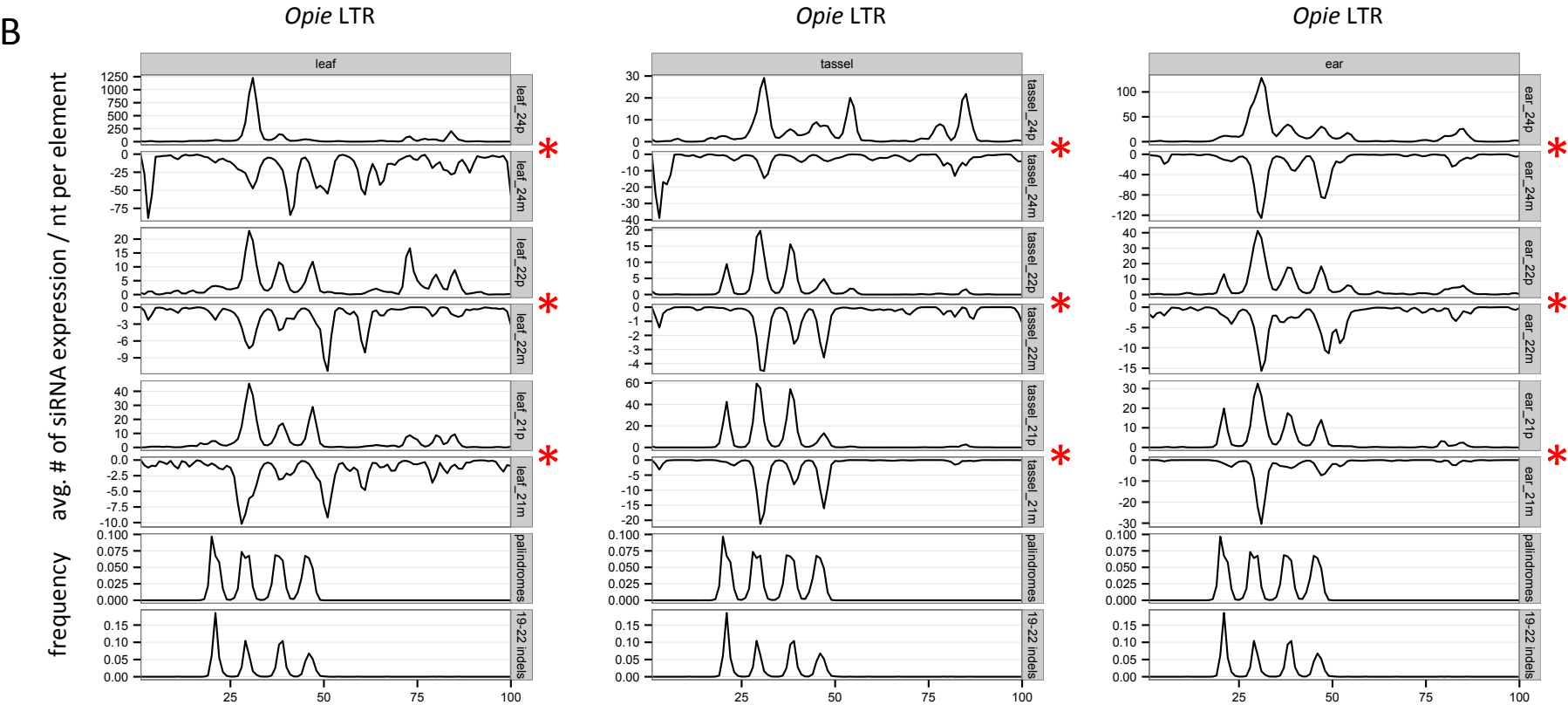
	Opie_24	Opie_22	Opie_21
leaf	2.11E-05	1.21E-06	5.07E-06
tassel	2.65E-03	2.64E-07	5.90E-09
ear	1.14E-05	3.41E-06	7.56E-08

B



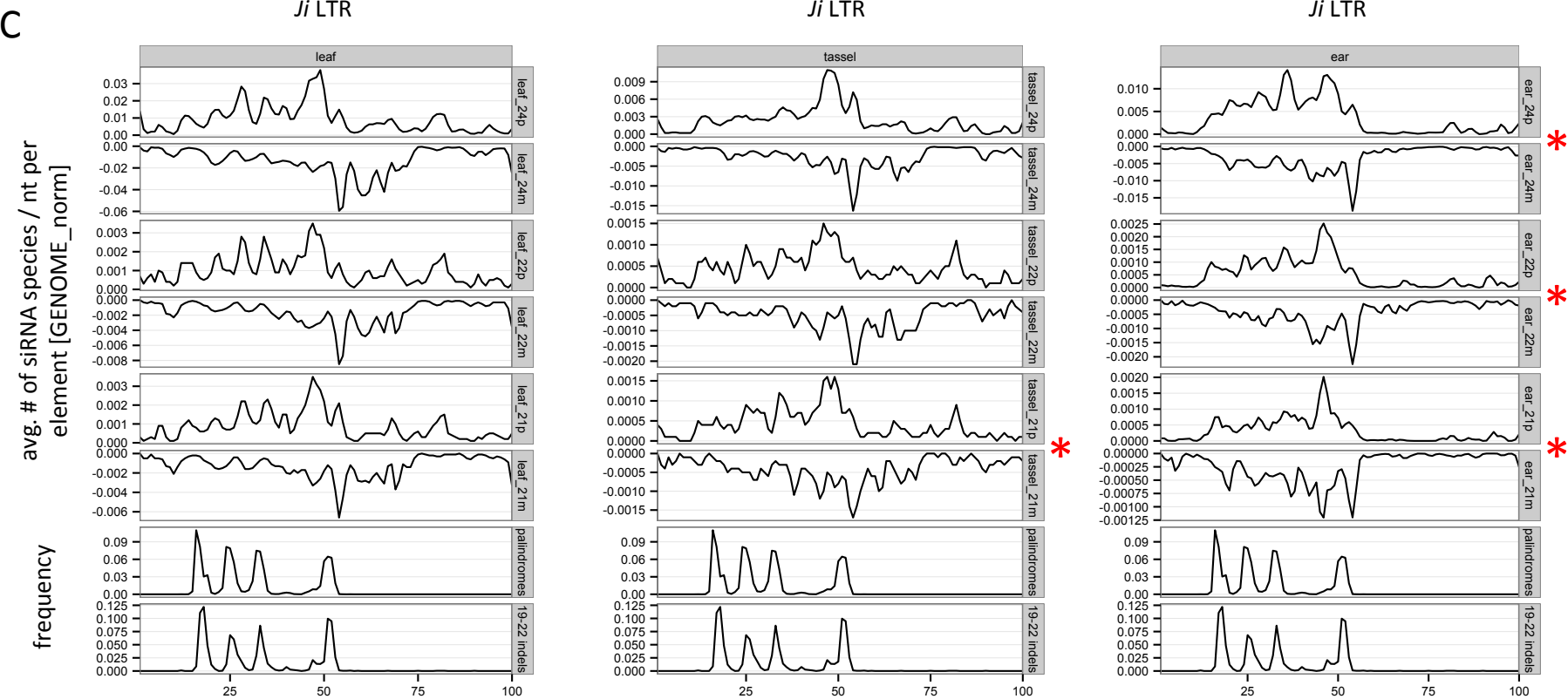
	Ji 24	Ji 22	Ji 21
leaf	1.54E-01	5.98E-01	2.85E-02
tassel	5.95E-01	1.98E-03	1.19E-07
ear	1.00E-05	1.85E-08	5.59E-09

Figure S6 (cont.)



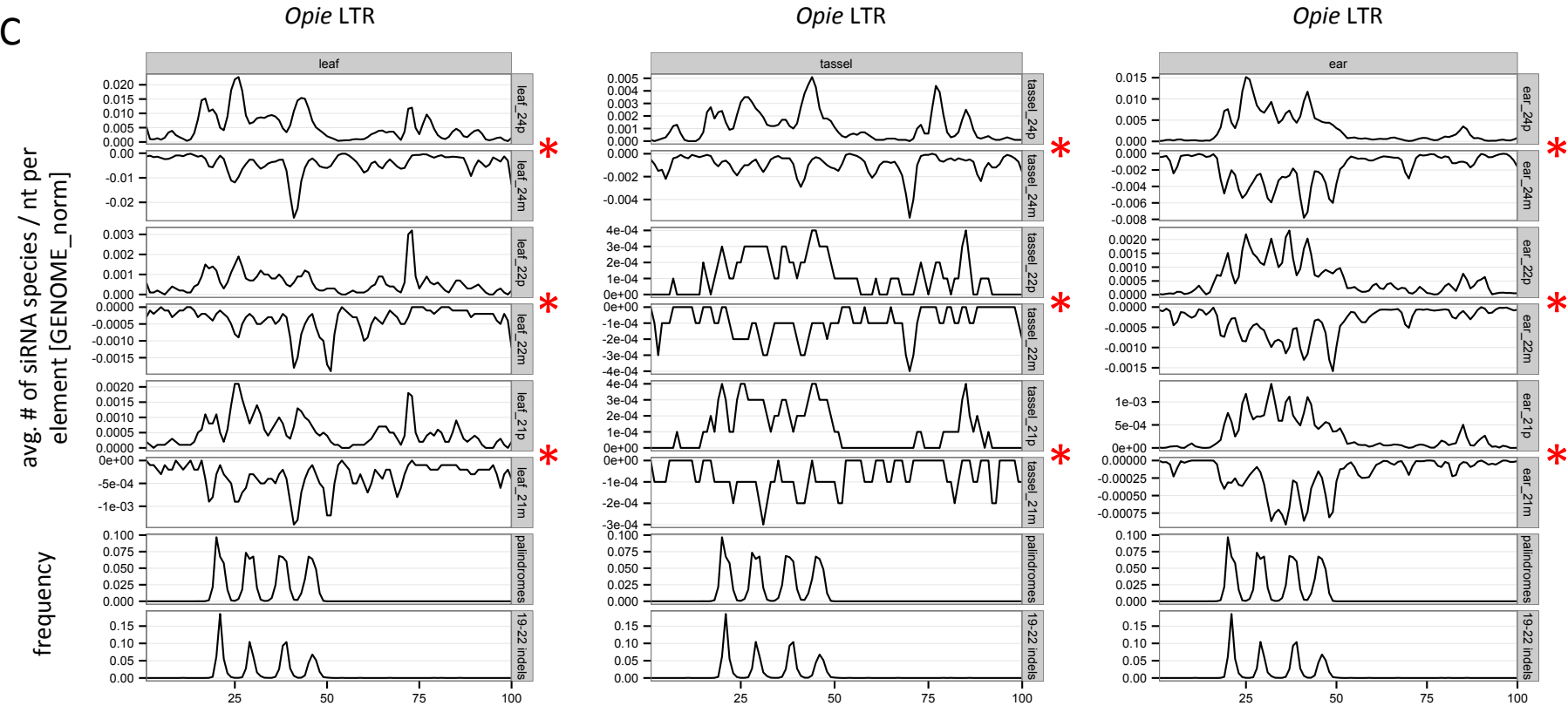
	Opie_24	Opie_22	Opie_21
leaf	2.07E-04	5.52E-06	2.39E-06
tassel	2.89E-03	1.72E-09	6.06E-10
ear	1.26E-06	1.69E-07	7.36E-09

Figure S6 (cont.)



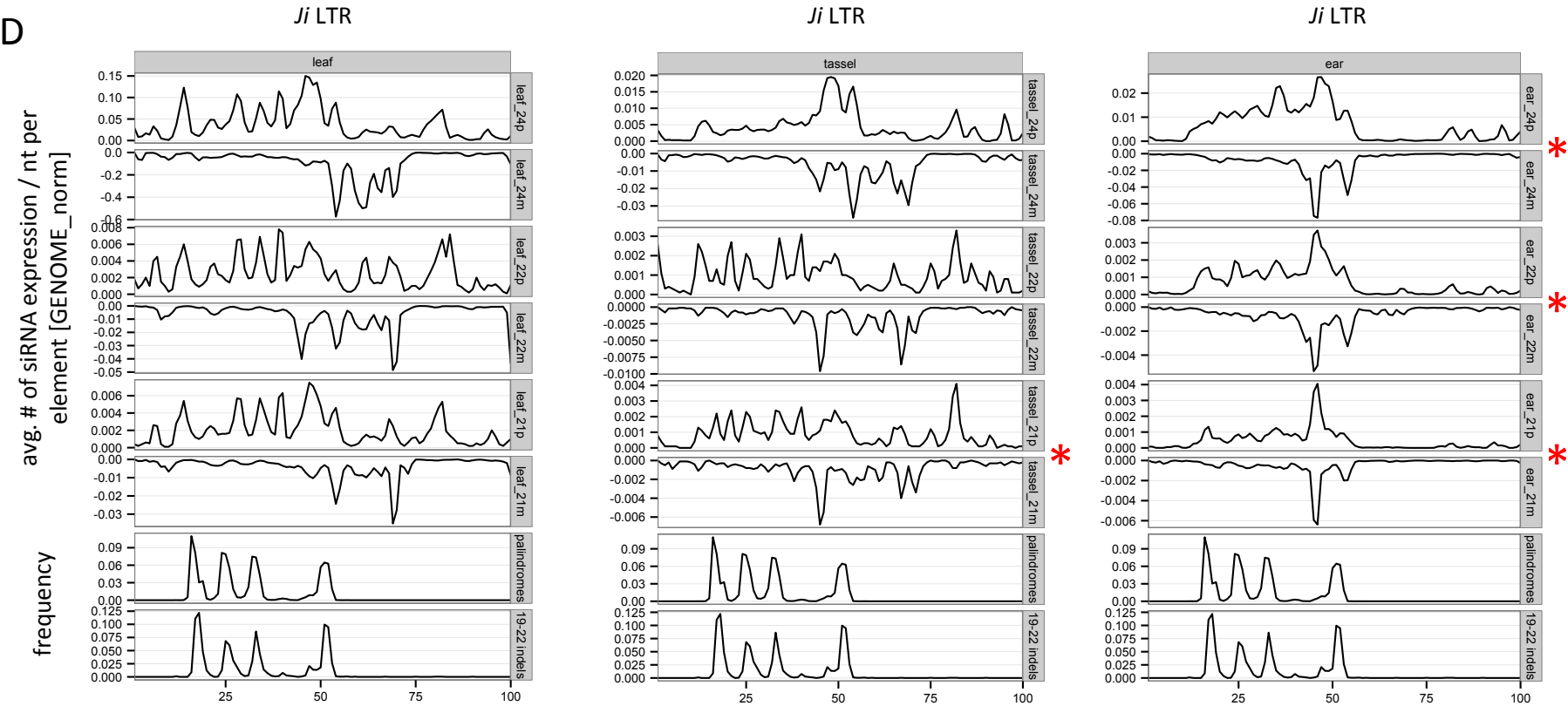
	Ji_24	Ji_22	Ji_21
leaf	3.12E-01	5.62E-01	2.63E-01
tassel	7.92E-02	9.92E-02	2.05E-03
ear	6.62E-04	3.16E-04	1.73E-04

Figure S6 (cont.)



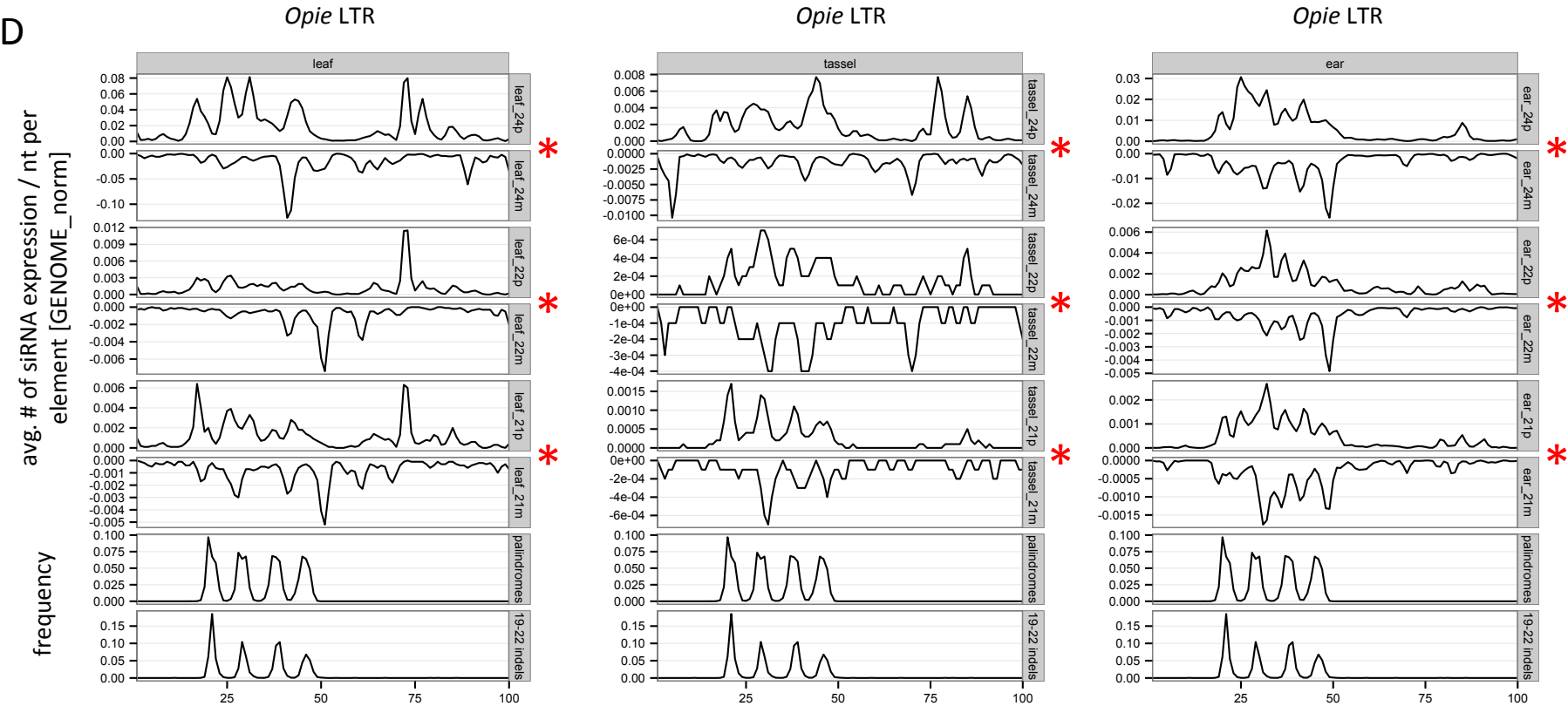
	Opie_24	Opie_22	Opie_21
leaf	2.69E-03	7.43E-03	8.34E-04
tassel	5.56E-04	2.53E-06	6.87E-08
ear	4.85E-06	1.09E-05	2.18E-06

Figure S6 (cont.)



	Ji_24	Ji_22	Ji_21
leaf	4.38E-01	8.74E-01	5.80E-01
tassel	4.13E-01	4.92E-01	1.99E-03
ear	6.84E-04	8.63E-05	1.40E-05

Figure S6 (cont.)



	Opie_24	Opie_22	Opie_21
leaf	1.72E-02	2.93E-02	2.24E-03
tassel	1.57E-03	2.66E-08	1.16E-10
ear	4.44E-06	3.73E-06	7.23E-07

Figure S7

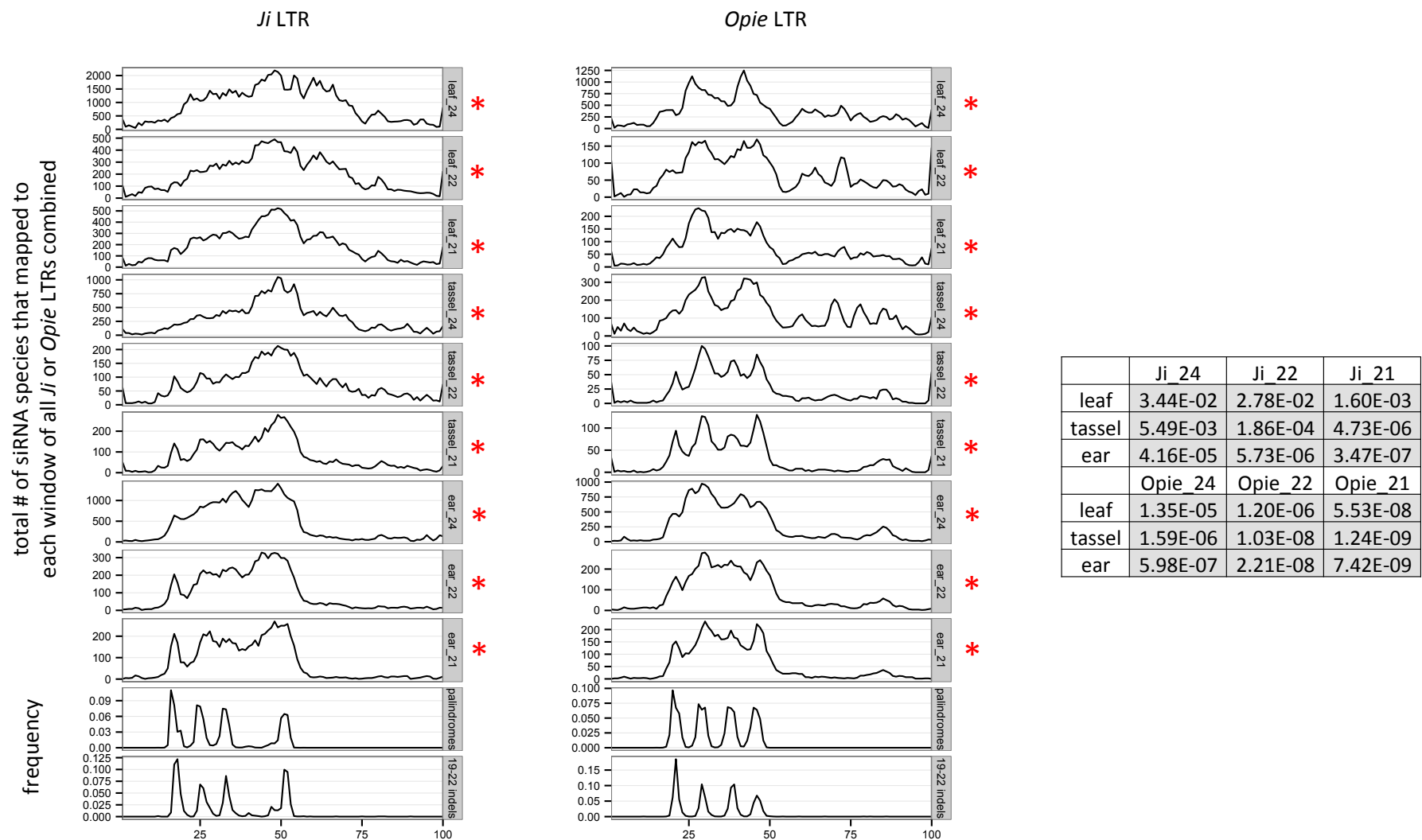


Figure S7. Total number of siRNA species that mapped to each window of all *Ji* or *Opie* LTRs. For these plots mapping information from the LTRs of 3,285 *Ji* (left) or 2,926 *Opie* (right) LTRs was combined. Each siRNA length and tissue was separately analyzed, while information for both strands was combined. Red asterisks indicate plots where the siRNA/palindrome overlap was statistically significant according to the Mann-Whitney U test (P values adjusted by Bonferroni correction). For the test, the four palindrome loci were combined, hence representing 16% of the LTR length, and tested for a significantly higher total number of siRNA species compared to the rest of the LTR. The bottom two plots of each panel show the distribution of palindromes and their 19-22nt indels. The table contains the P values of the Mann-Whitney U test for each combination of family, siRNA length and tissue. Cells highlighted in grey depict the statistically significant P values.

Figure S8

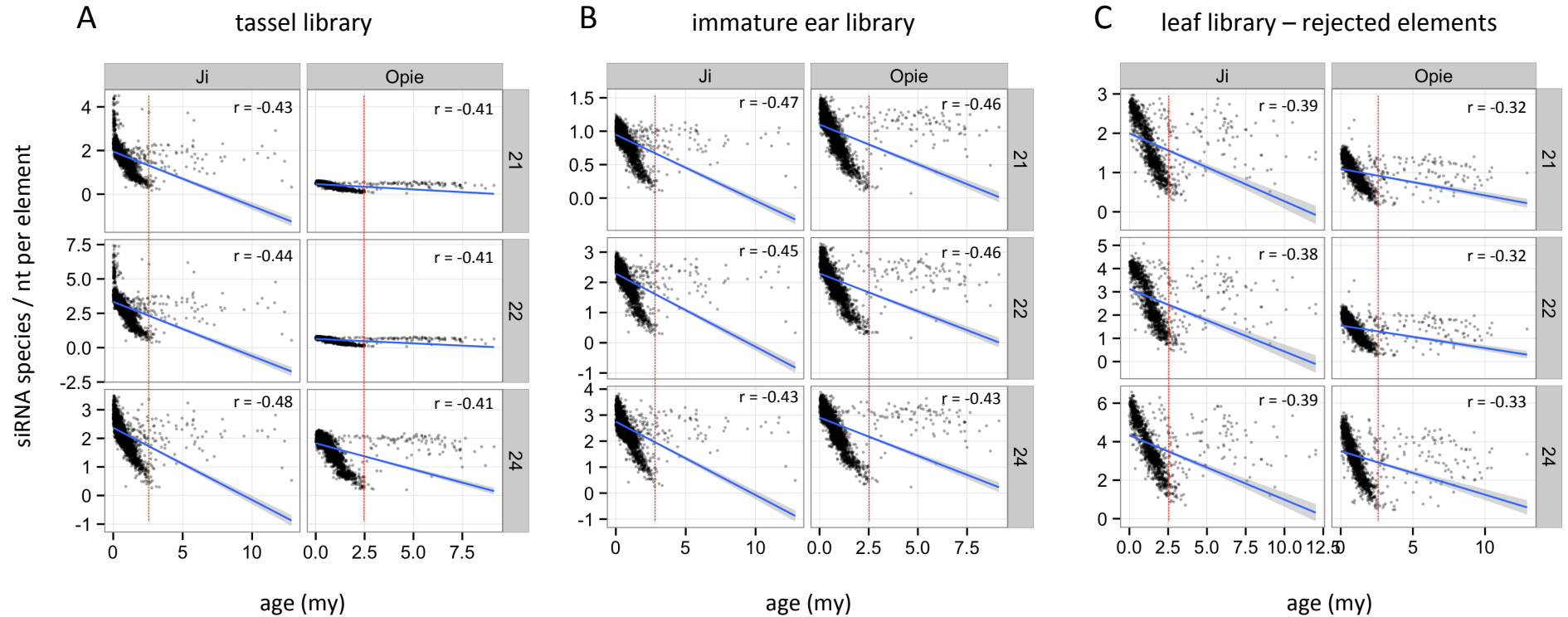


Figure S8. Insertion age and siRNA targeting for *Ji* and *Opie* elements using different datasets and libraries. Relationship between age and siRNA species (calculated per nucleotide of the full-length sequence) for all siRNA lengths using the tassell (A) or immature ear (B) libraries, or the leaf library but analysing a subset of elements that were rejected along the filtering pipeline, because they were outside the typical family ranges or contained insertions from other TEs (C). These 1,291 *Ji* and 1,561 *Opie* elements did not contain >5 'N' nucleotides in their sequences. The grey shading around each regression line represents the 0.95 confidence interval level, while the Pearson r is shown for each plot (P values were $<10^{-20}$ in all cases). The dotted red lines indicate the ~2.0-2.5my point, after which elements did not exhibit decreased targeting.