

Figure S9

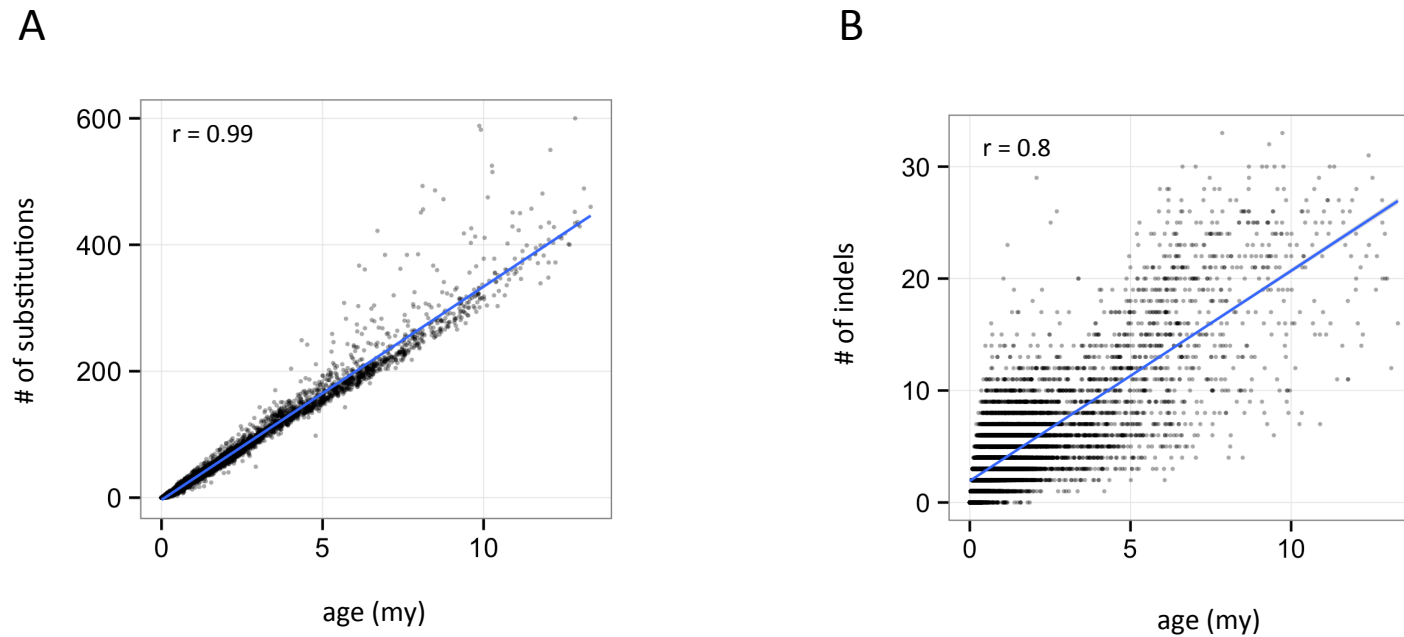


Figure S9. Number of indels and nucleotide substitutions in the LTR pair of each Sirevirus element as a function of age. The total sample of 6,456 elements was used in this analysis. The grey shading around each regression line represents the 0.95 confidence interval level, while the Pearson r is shown for each plot (P values were $<10^{-20}$ in both cases). my, million years.

Figure S10

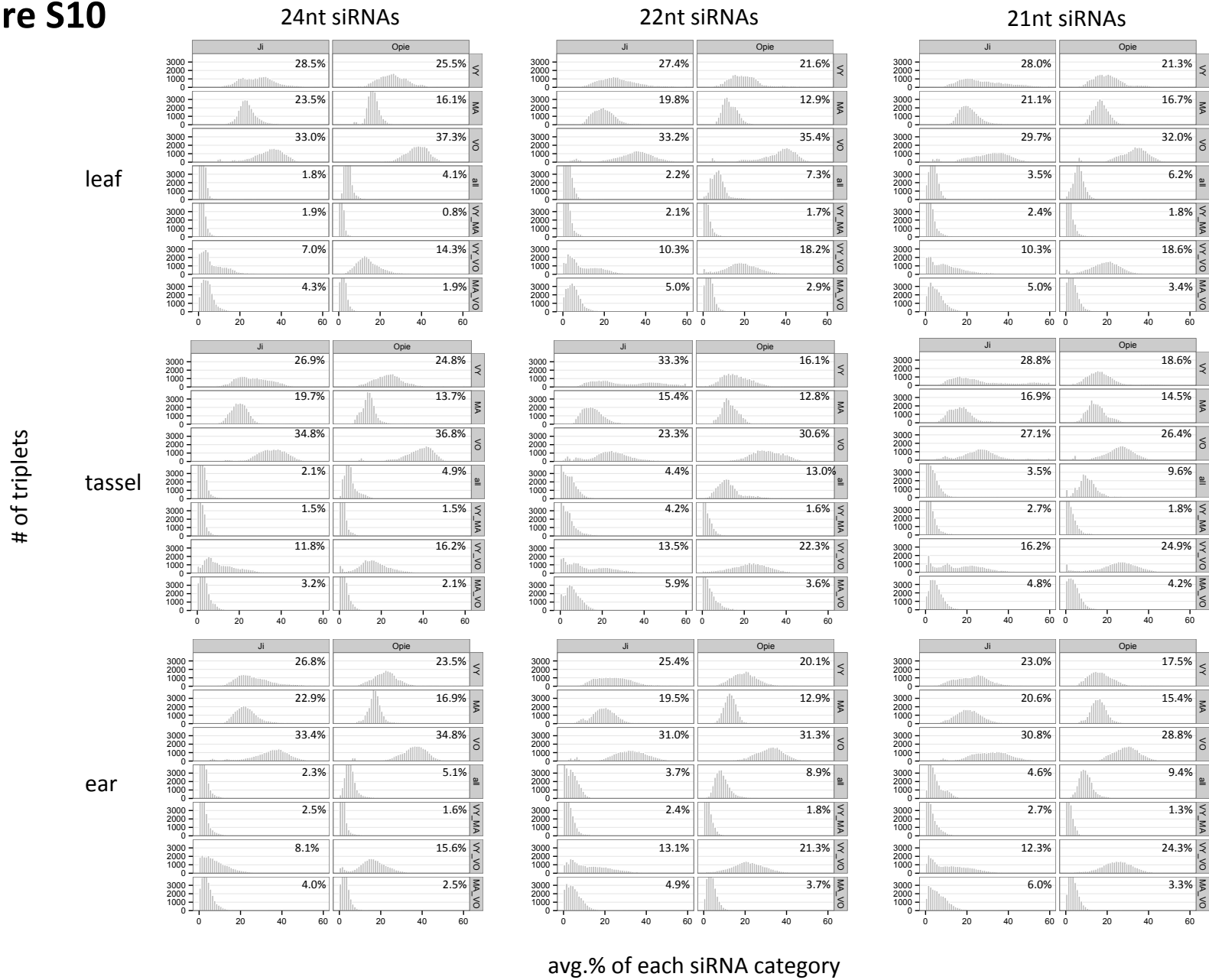


Figure S10. siRNA species cross-mapping to the LTRs of 27,000 *Ji* or *Opie* triplets. Each siRNA was classified across seven types, based on the element(s) of the triplet it mapped to. The average percentage of each type is shown. The y axis is cut-off at 4,000. VY, very young; MA, middle-aged; VO, very old.

Figure S11

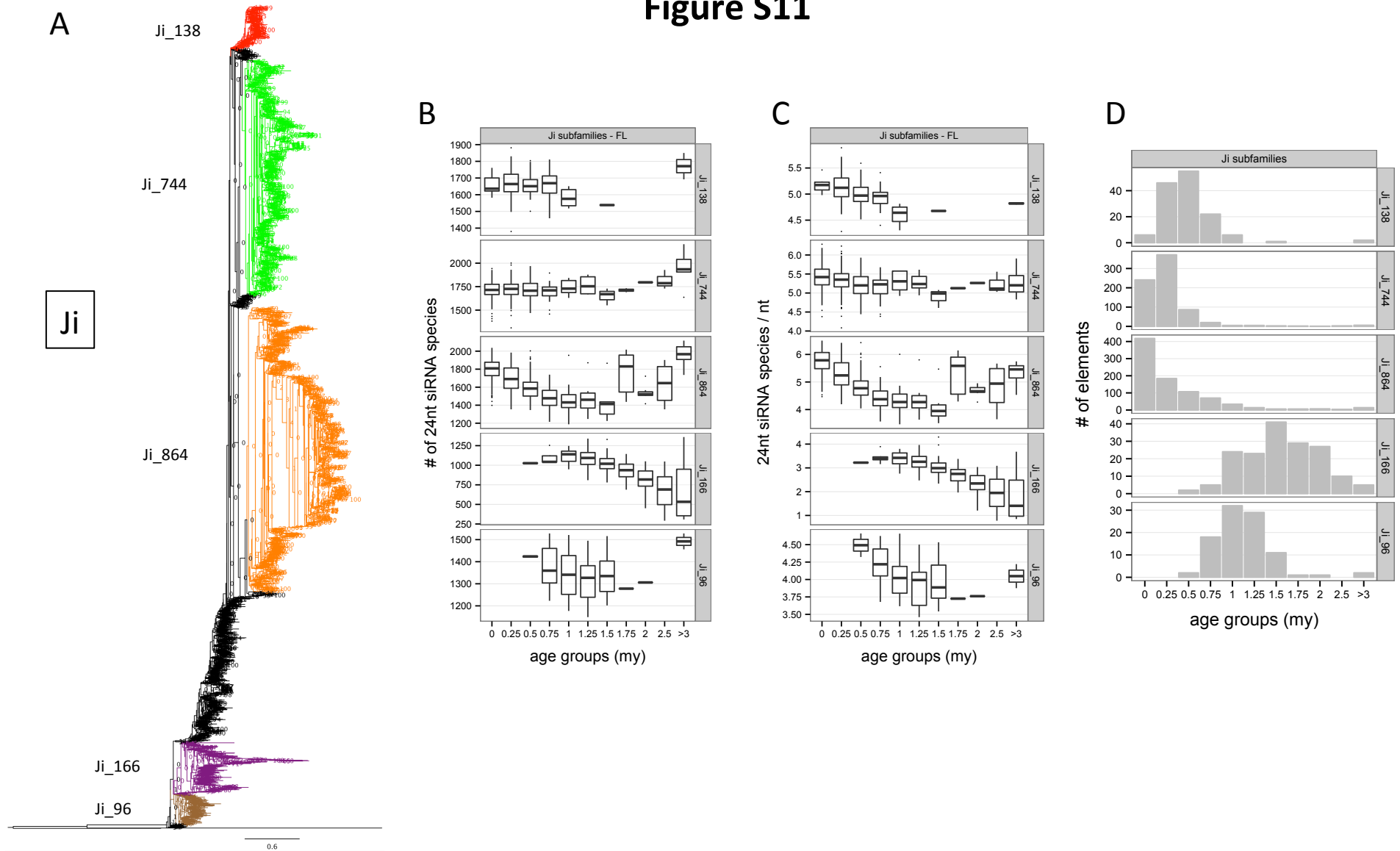


Figure S11. *Ji* and *Opie* phylogenies. **(A)** Maximum likelihood phylogenetic tree of the *Ji* and *Opie* elements based on the core domain (Pfam model PF07727) of their reverse transcriptase gene. The phylogenetic analysis was performed using the RAxML package with default parameters (Stamatakis 2014). The bootstrap confidence levels are shown at the nodes of the tree. Color-coded clades participated in the analysis of the relationship between age and siRNA targeting. The name of each clade also indicates its size (e.g. *Ji*_138). **(B)** Number of 24nt siRNA species of the leaf library, or **(C)** their per nucleotide average that mapped to the full-length sequences across age groups of individual clades. In (C), the pattern remained consistent, despite averaging mapping per nucleotide of each element to control for changes in element's length. **(D)** Population distribution across age groups for *Ji* and *Opie* clades. my, million years.

Figure S11 (cont.)

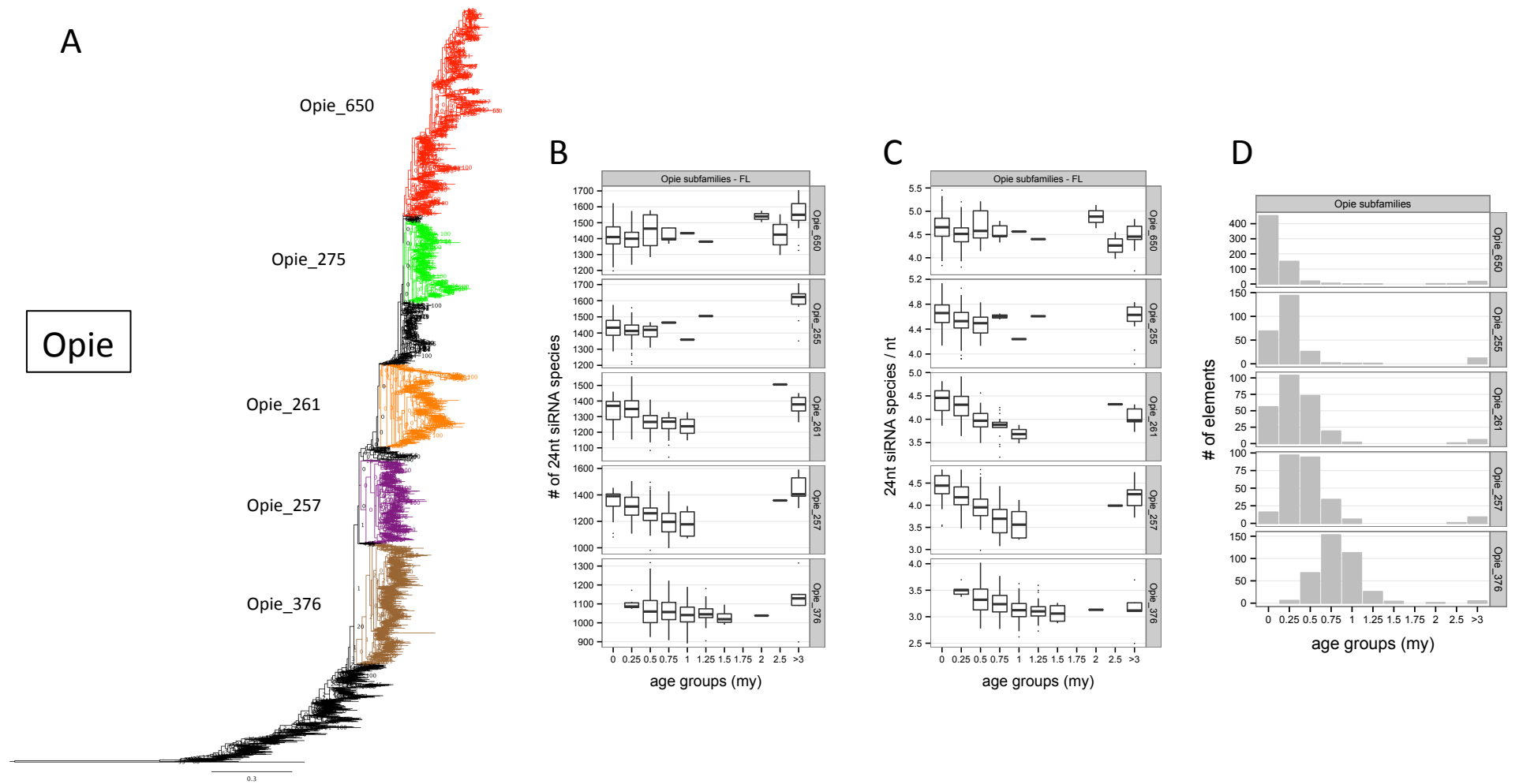


Figure S12

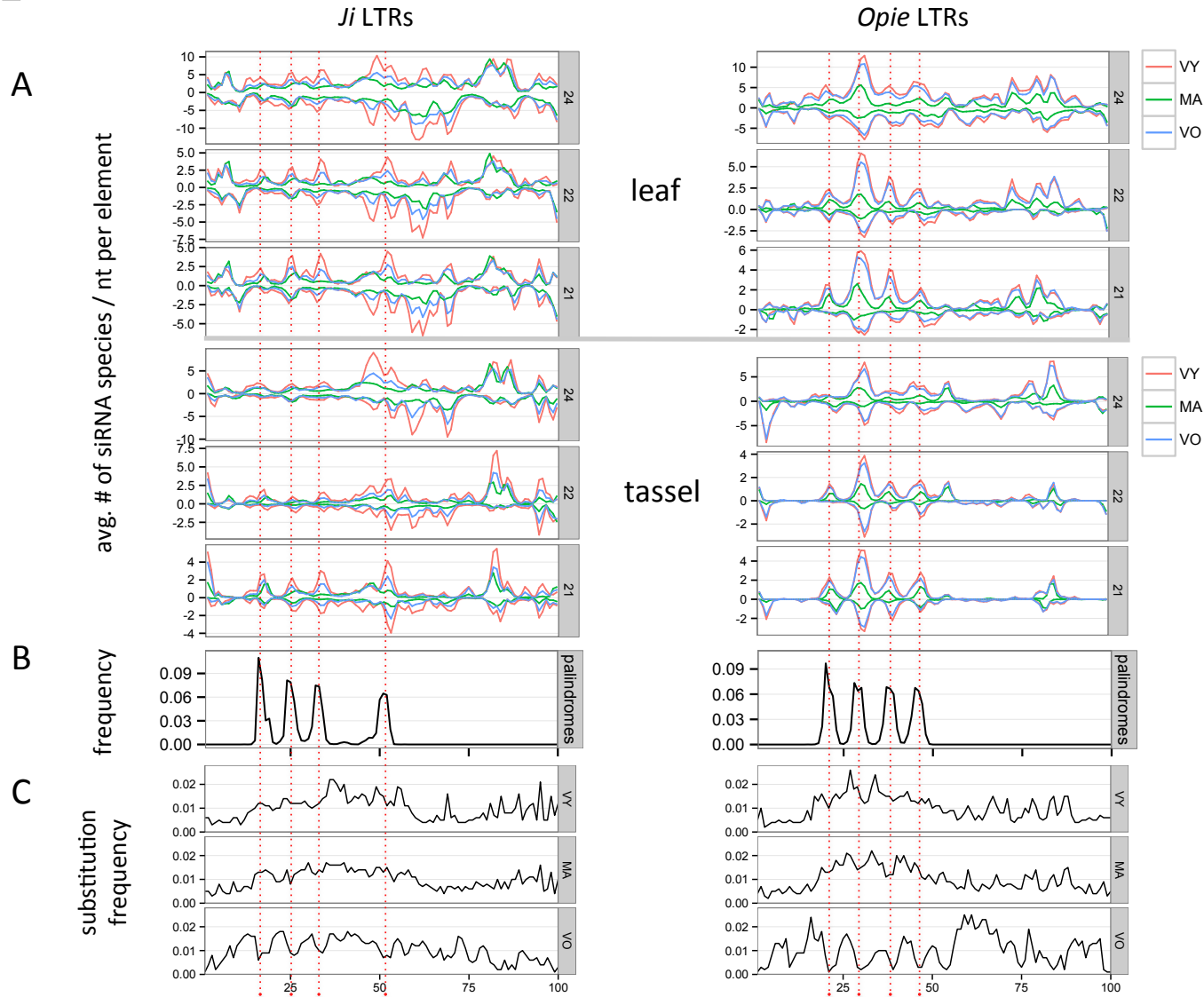


Figure S12. siRNA targeting and conservation of the palindrome-rich region as a function of age. Each LTR and INT domain of *Ji* and *Opie* elements was first split in 100 equally-sized windows of ~13nt and ~67nt length respectively. **(A)** Number of siRNA species of the leaf and tassel libraries per nucleotide of the sense (positive x-axis) and antisense (negative x-axis) strands of the LTRs of very young (VY), middle-aged (MA) and very old (VO) elements (see Table S4 for their population sizes). The average of each window is depicted instead of a boxplot that contains all measurements. **(B)** Distribution of palindromes. **(C)** Distribution of substitutions identified during LTR pair alignment. **(D)** As in (A) but analyzing the INT domain of *Ji* and *Opie*.

Figure S12 (cont.)

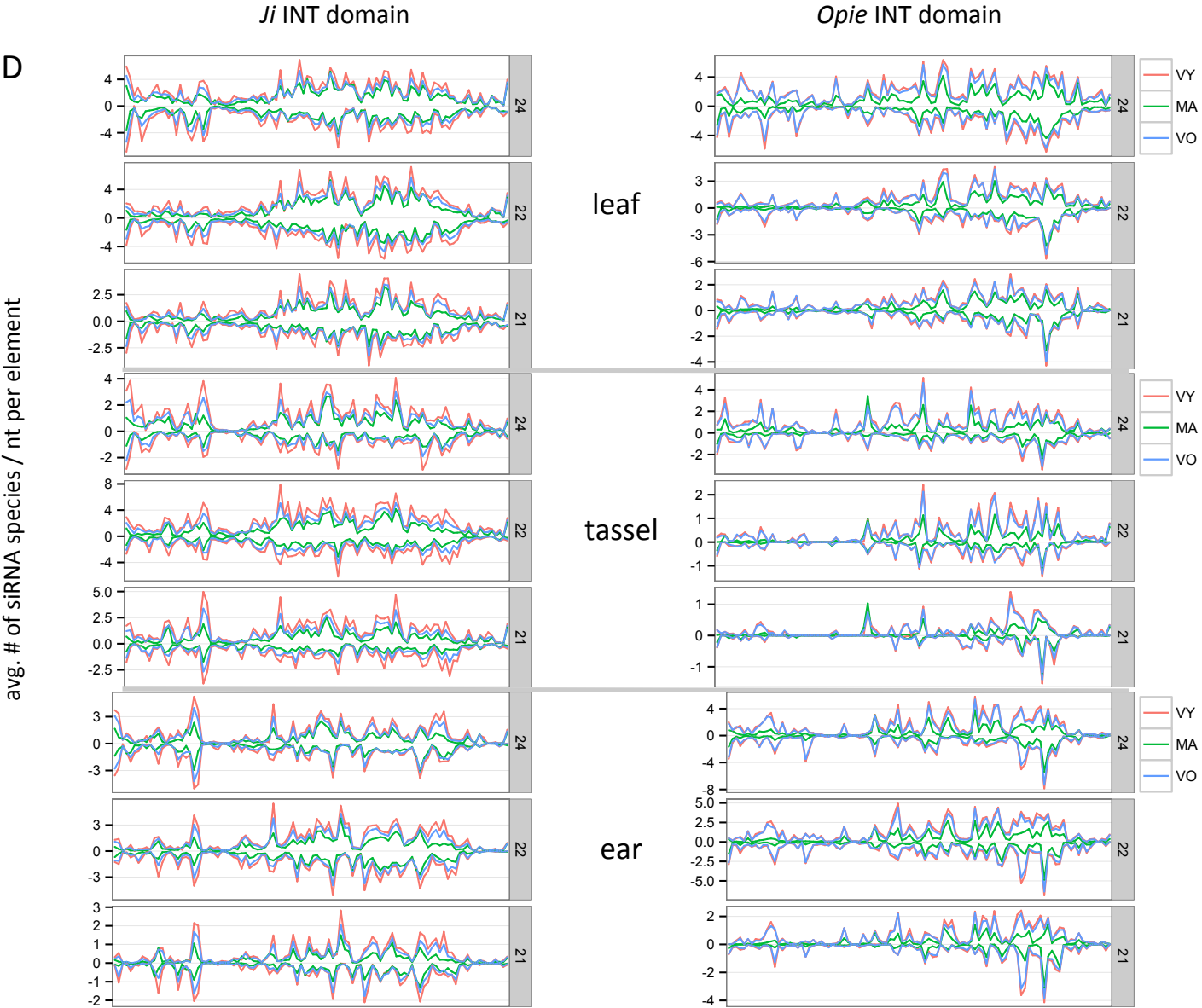


Figure S13

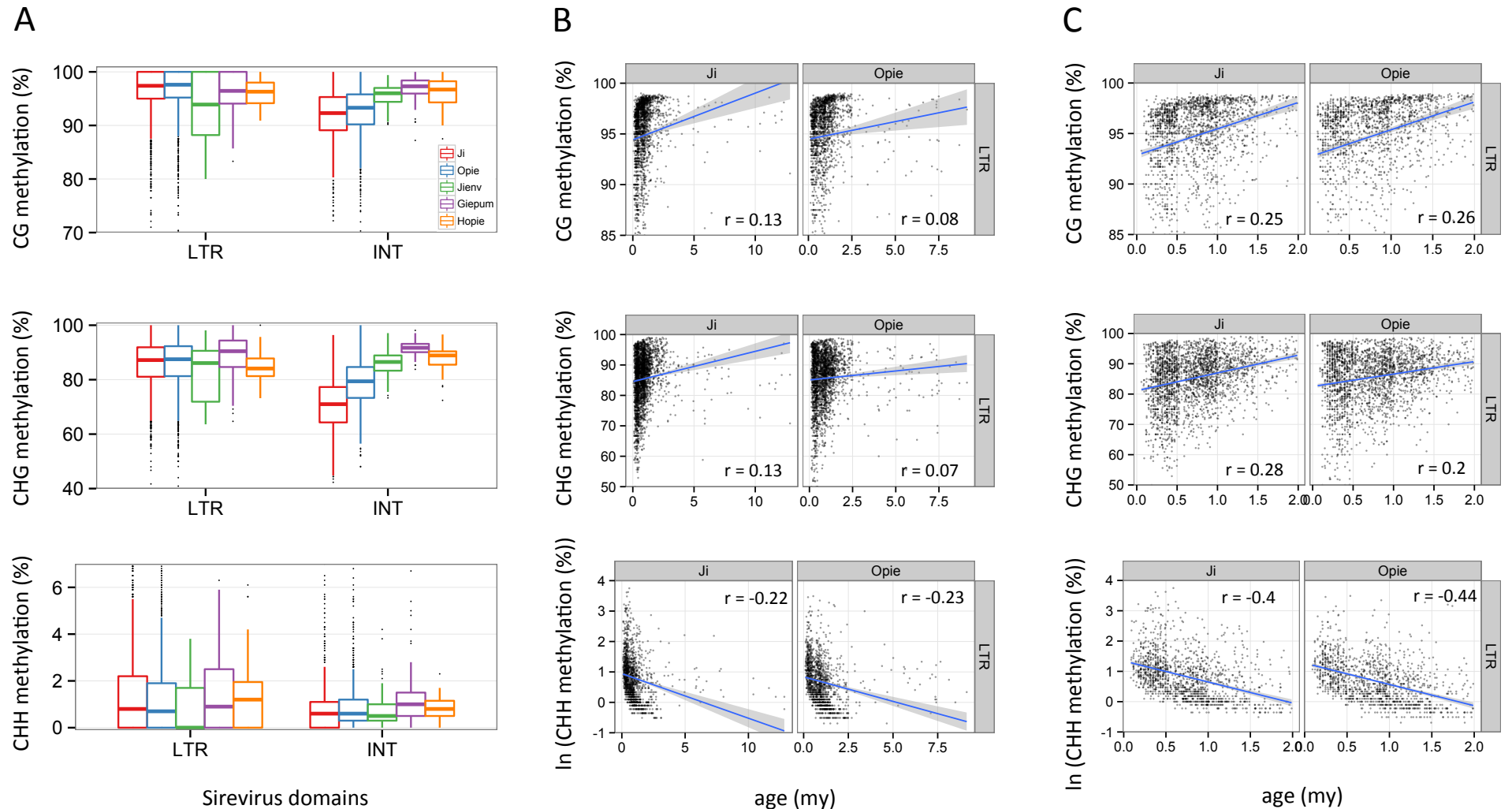


Figure S13. Methylation patterns of Sireviruses. **(A)** CG, CHG and CHH methylation levels for the LTRs and INT domain of each family. **(B, C)** Relationship between methylation levels and age for the *Ji* and *Opie* LTRs. For these plots, LTRs with maximum CG or CHG, or minimum CHH methylation were excluded (see Table S8). For (C) only elements between 0-2my were examined. 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 most cases. my, million years.

Figure S14

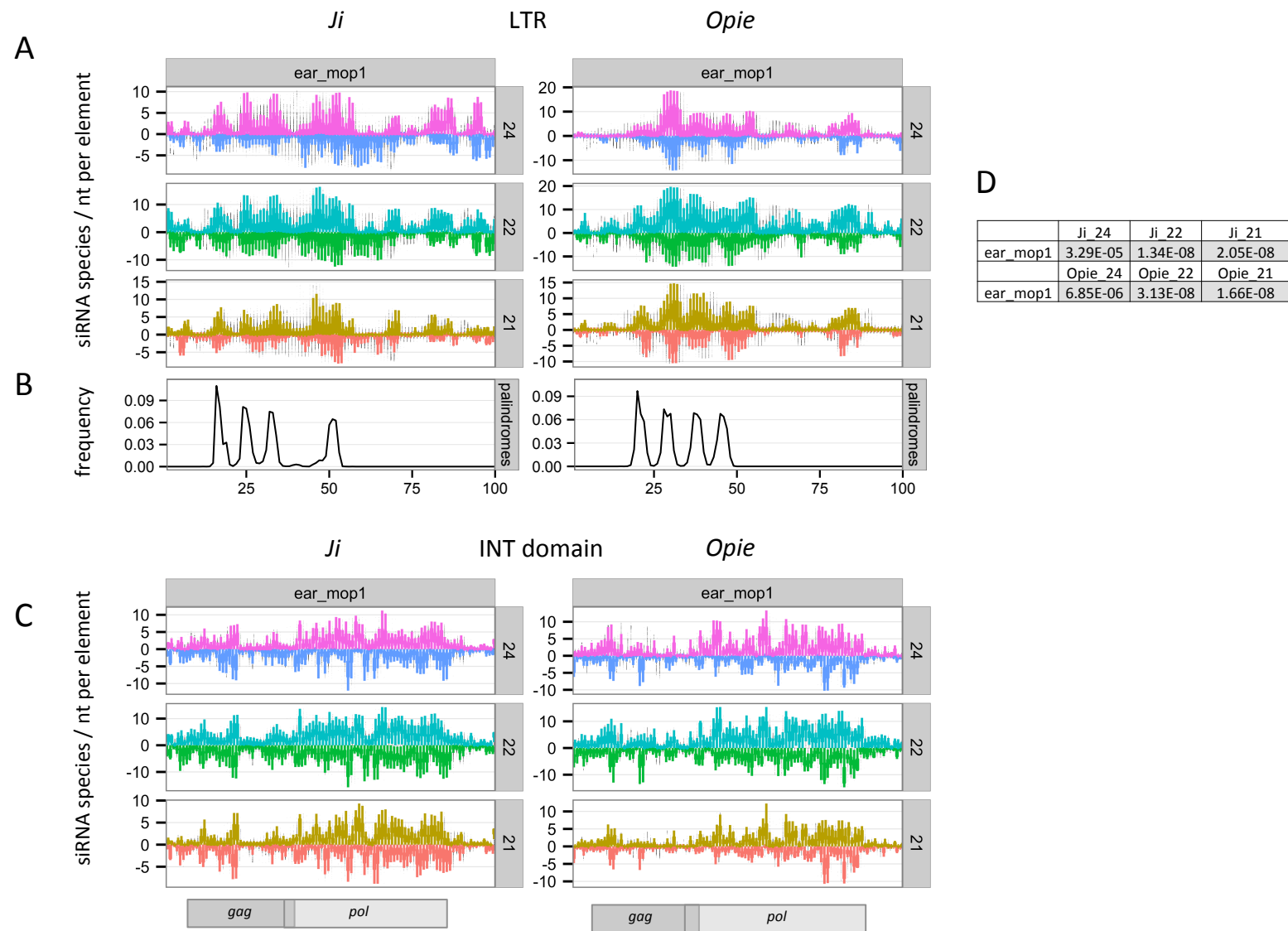


Figure S14. siRNA mapping patterns of ear *mop1* library to *Ji* and *Opie*. Each LTR and INT domain was first split in 100 equally-sized windows of ~13nt and ~67nt length respectively. **(A)** Number of siRNA species per nucleotide along the sense (positive x-axis) and antisense (negative x-axis) strands of the LTRs of 3,285 and 2,926 *Ji* and *Opie* elements respectively (visualized with a boxplot for each window). **(B)** Distribution of palindromes. **(C)** As in (A) but showing the INT domain. The approximate positions of the *gag* and *pol* genes are shown. **(D)** *P* values of the Mann-Whitney U test for each combination of family and siRNA length. Cells highlighted in grey depict the statistically significant *P* values.

Figure S15

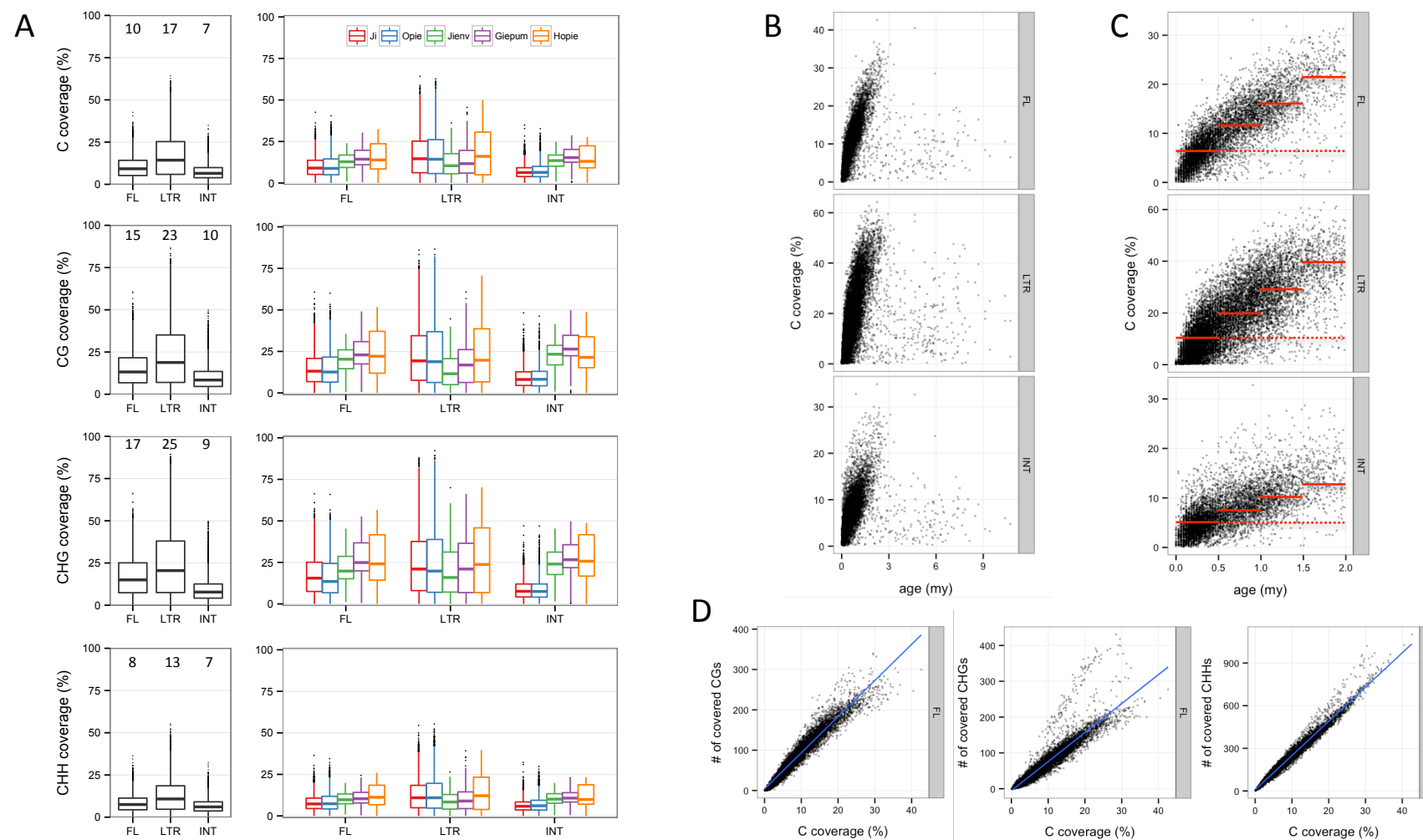


Figure S15. Bisulfite sequencing coverage and filtering cutoffs. **(A)** Coverage (≥ 2 -fold) of the full-length, LTRs or INT domain of all 6,456 elements (left) or of each family separately (right). The y-axis represents the proportion of cytosines that were covered by ≥ 2 reads. The values within the plots indicate the average coverage of each domain. **(B)** Relationship between age and the proportion of coverage for all cytosine contexts combined. **(C)** Like (B), but showing the age-adjusted coverage cutoffs in 0.5my windows (red lines). Dotted red lines indicate the cutoffs for elements older than 2my, and also an age-independent horizontal filter applied across the entire population. **(D)** Relationship between the proportion of coverage for all cytosine contexts combined and the number of covered CG, CHG or CHH sites for full-length elements.