# Supplementary Methods

## Analyzing symmetric/anti-symmetric components of N-to-C SSR distribution

To study the asymmetry of SSR density along N–>C position, we performed a simple decomposition of each curve into its purely symmetric and purely asymmetric parts. For a given function $f(x)$, where $0 \le x \le 1$ is the relative position along the coding sequence, we defined the symmetric and anti-symmetric parts:

$$f_{sym}(x) = \frac{1}{2}\left(f(x) + f(1-x)\right) \qquad f_{anti}(x) = \frac{1}{2}\left(f(x) - f(1-x)\right)$$

These two functions are respectively symmetric and anti-symmetric about the middle of the coding sequence (x = 0.5), and the decomposition satisfied $f(x) = f_{sym}(x) + f_{anti}(x)$. In many species that exhibit anti-symmetry (see Supplementary Figure 2B for example), the N-terminal is enriched in repeats relative to the C-terminal. In those cases, we consistently find that the anti-symmetric portion of the primary sequence signal (red curves) is significantly smaller in magnitude than the anti-symmetric portion of the observed data (blue).

## Analyzing SSR distribution using a more stringent length cutoff

Since it is informative to understand the selective pressure of *long SSRs*, we performed similar analysis in SSR frequency (as in Figure 6 and Supplementary figure 2), except using a more stringent length cutoff as following:

| | Monomeric SSRs | Dimeric SSRs | Trimeric SSRs | Tetrameric SSRs |
|---|---|---|---|---|
| Normal length cutoff | >= 6bp | >= 6bp | >= 9bp | >= 12bp |
| Stringent length cutoff | >= 8bp | >= 8bp | >= 12bp | >= 16bp |

We did not change the cutoff of k-mer SSRs with k >4, since these SSRs are much fewer in numbers (<1% of total SSR) and contribute little to overall statistics.

After imposing stringent cutoffs, only 5%~10% of SSRs identified using the normal cutoffs pass these criteria. The resulting analysis using stringent cutoffs are shown in Supplementary Figures 3 and 4, and discussed below.

Overall SSR avoidance/enrichment:

- Among 20 species that exhibit avoidance using original cutoffs, after increasing cutoffs, all of them still exhibit significant avoidance.

- Among the 11 species that originally exhibited overall SSR enrichment, after increasing the cutoffs, 5 species (Rp, Bcn, Bm, Bp, Xc) remain at significant enrichment, 2 species (Nm, Se) change from enrichment to no significant enrichment/avoidance, and 4 species (Cbu, Pa, Cj, Hp) change from enrichment to avoidance.

- Among 10 species that originally showed no enrichment or avoidance, with the increased cutoffs, 6 species (Ype, Yps, Ft, Lp, Pm, Mm) exhibit avoidance while 4 species exhibit no change.

Overall, increasing the SSR cutoff to the more stringent level tends to yield more avoidance, which supports the argument that long SSRs are more strongly selected against. Yet, after eliminating 90% of short SSRs, 30 species did not change their overall tendency, and 5 species still exhibit significant enrichment of total SSRs. This indicates that in many species, the overall tendency of enrichment or avoidance is not sensitive to specific parameter settings in the analysis.

Type-specific SSR enrichment/avoidance

Since total SSR enrichment/avoidance is a composite effect of individual SSR types, we tested whether the enrichment/avoidance in different SSR types is robust after changing length cutoffs. To see this, we focus on monomeric and dimeric SSRs, for which we have the best statistics. Among 10 types of monomeric/dimeric SSRs in 42 species, we have 420 type-specific bins, each represent type- and species-specific SSR enrichment/avoidance (see Figure 5 for the analysis result in original cutoff).

We checked the original/shuffled SSR ratio (R) to see if this parameter is robust. Among 420 bins, 80% remain in the same enrichment/avoidance tendency after cutoff change. Of the 20% that change tendency, most do not exhibit a significant change. Detailed analysis shows that only 9 in 420 bins (2.2%) change from significant enrichment ($R>1.25$) to significant avoidance ($R<0.75$), and only 7 in 420 (1.7%) change from significant avoidance ($R<0.75$) to significant enrichment ($R>1.25$).

Therefore, we conclude that, after changing cutoffs such that 90% of short SSRs are eliminated from statistics, the type-specific SSR enrichment/avoidance is generally robust.

Asymmetric N-terminal enrichment of SSRs

We re-analyzed the asymmetric distribution of SSRs from N->C terminals, giving the decomposition into symmetric and asymmetric parts in the new Supplementary Figure 4. In many species, the asymmetric portion, shown in Fig S4b, still exhibits the same trends we had previously pointed out using the original cutoffs