Document S2

Bootstrap analysis

Medium level correlation could be partially due to the fact that only the already identified determinants are correlated with the AEV values and the lack of hitherto unidentified determinants should weaken the observed relationship. Nevertheless, even based on a simple visual observation positional distribution of AEV and NPD appeared to be non-independent. In order to demonstrate that AEV and NPD values are distributed non-independently we performed a different statistical test, a bootstrap analysis.

We summed up the AEV values of the 40 (out of 96) tRNA positions that harbor at least one published determinant. This gave a cumulative value of 345.55. We aimed to estimate the probability of getting a cumulative AEV equal to or higher than 345.55 by chance. Therefore we randomly picked 40 positions and calculated the cumulative AEV and we performed this one million times. The results are plotted as probability density distribution figure below. Importantly, none of the random sets from the analysis yielded equal or higher cumulative AEV. This means that the probability of randomly picking a set of 40 residues (out of 96) that would provide a cumulative AEV reaching or exceeding the cumulative AEV of the set of positions of known determinants is less than 10-6. The bootstrap analysis generated probability density distribution had a mean value of 258 and standard deviation value of 17.1. A calculation based on these two data and the cumulative distribution function suggests that the probability of reaching or exceeding the AEV value of 345.55 is 3.54 x 10-7, which is negligible.

Thus, regarding their AEV values, the set of positions harboring already published determinants (positions with a nonzero NPD) is definitely not random: those positions typically coincide with high AEV values. This suggested that high AEV positions accompanied with zero or low NPD could locate positions where potential hidden determinants exist.

Results of a bootstrap analysis assessing correlation of AEV and NPD values

There are 40 out of 96 tRNA positions where determinants had been experimentally identified in bacteria (i.e. NPD is not zero). Summing up the AEV values of these 40 positions provides a cumulative AEV of 343.55 indicated by an arrow on the X axis. We assessed the probability of reaching or exceeding this cumulative AEV value by randomly picking 40 positions and calculating the cumulative AEV. The histogram corresponds to one million such random selections and calculations and illustrates the probability density plot of the results. It is readily apparent that high AEV values are not randomly distributed among positions as the set of 40 positions with nonzero NPD corresponds to a very high cumulative AEV. It is very improbable to obtain the same or higher cumulative AEV by random picking of 40 positions. As explained in the Results the probability of that is estimated to be 3.54 x 10^{-7}, which is negligible. Details of the analysis are provided in the Results section.