

## **tRNomics: Genomic Organization and Processing Patterns of tRNAs**

### **Abstract**

Surprisingly little is known about the organization and distribution of tRNAs and tRNA-related sequences on a genome-wide scale. While tRNA complements are usually reported in passing as part of genome annotation efforts, and peculiar features such as the tandem arrangements of tRNAs in *Entamoeba histolytica* have been described in some detail, comparative studies are rare. We therefore set out to systematically survey the genomic arrangement of tRNAs in a wide range of eukaryotes to identify common patterns and taxon-specific peculiarities. We found that tRNA complements evolve rapidly and that tRNA locations are subject to rapid turnover. At the phylum level, distributions of tRNA numbers are very broad, with standard deviations on the order of the mean. Even within fairly closely related species, we observe dramatic changes in local organization. Consistent with this variability, syntenic conservation of tRNAs is also poor in general, with turn-over rates comparable to those of unconstrained sequence elements. We conclude that the genomic organization of tRNAs shows complex, lineage-specific patterns characterized by extensive variability, and that this variability is in striking contrast to the extreme levels of sequence-conservation of the tRNA genes themselves. Our comprehensive analysis of eukaryotic tRNA distributions provides a basis for further studies into the interplay between tRNA gene arrangements and genome organization in general.

Secondly, we focused on the investigation of small non-coding RNAs (ncRNAs) from whole transcriptome data. Since ncRNAs constitute a significant part of the transcriptome, we explore this data to detect and classify patterns derived from transcriptome-associated loci. We selected three distinct ncRNA classes: microRNAs, snoRNAs and tRNAs, all of which undergo maturation processes that lead to the production of shorter RNAs. After mapping the sequences to the reference genome, specific patterns of short reads were observed. These read patterns appeared to reflect RNA processing and, if so, should specify the RNA transcripts from which they are derived. In order to investigate whether the short read patterns carry information on the particular ncRNA class from which they originate, we performed a random forest classification on the three distinct ncRNA classes listed above. Then, after exploring the potential classification of general groups of ncRNAs, we focused on the identification of small RNA fragments derived from tRNAs. After mapping transcriptome sequence data to reference genomes, we searched for specific short read patterns reflecting tRNA processing. In this context, we devised a common tRNA coordinate system based on conservation and secondary structure information that allows vector representation of processing products and thus comparison of different tRNAs by anticodon and amino acid.

We report patterns of tRNA processing that seem to be conserved across species. Though the mechanisms and functional implications underlying these patterns remain to be clarified, our analysis suggests that each type of tRNA exhibits a specific pattern and thus appears to undergo a characteristic maturation process.