Development of a novel approach for 100x faster identification of all five gene's structural classes for the non-coding, catalytic RNase P

Background

- RNase P is an essential, catalytic non-coding RNA (ncRNA) that is ubiquitous in all three kingdoms (Bacteria, Archaea, and Eukaryotes).
- This ncRNA is challenging to annotate due to both variable genomic sequences and diverse secondary structures.

Approach

- To create a user-friendly graphical user interface for the rapid identification of RNase P from genomic data.

Outcome

- The newly developed algorithm can identify RNase P RNA 100x faster than the leading software. P Finder is available at https://github.com/JChristopherEllis/P-Finder.
- Five distinct RNase P structure types are observed in prokaryotes and the new algorithm is the first that can identify all of them, including the rare C-type and T-type RNase Ps.
- The software's taxonomic assignment tool facilitates synthetic microbiome studies.

Significance

- This rapid and accurate software tool is enabling the detection of new RNase Ps and facilitating the taxonomic characterization of genomes and microbiomes.
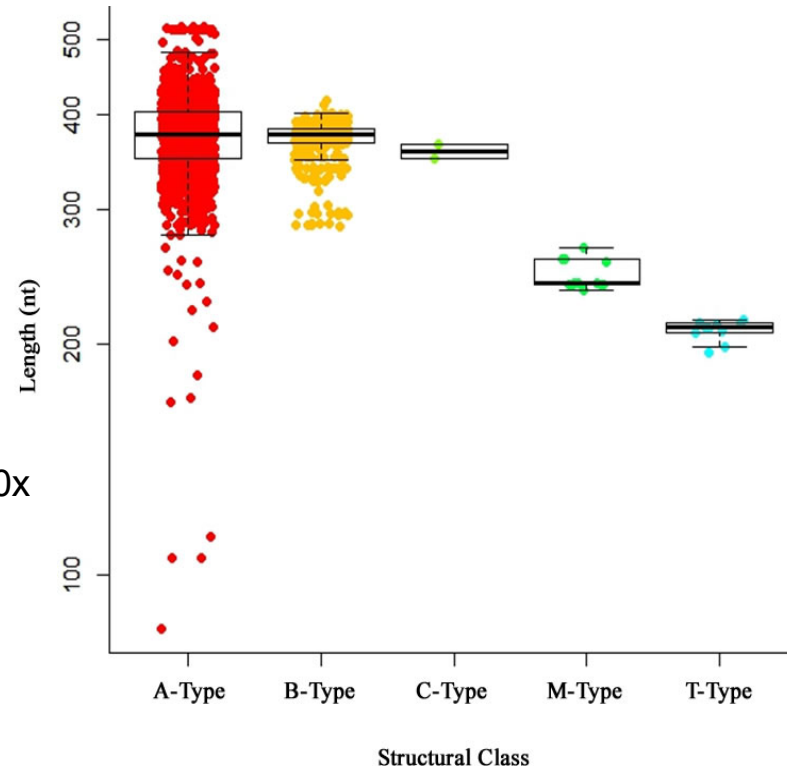
Distribution of RNase P RNA structural classes by type and length. RNase P RNA has a broad diversity of sequence length and structure. A-type RNase P RNAs are the most common structural class. The minimal T-type, C-type, and M-type are uncommon with only 28 organisms identified to date containing one of these structural classes.

J. Christopher Ellis, "P finder: genomic and metagenomic annotation of RNase P RNA gene (rnpB)," *BMC Genomics* 21:334. doi: 10.1186/s12864-020-6615-z