1. Introduction

- Recently, non-protein-coding RNAs (ncRNAs) have moved from a biochemical curiosity to a main research topic in molecular biology.
- High-throughput transcriptome analyses have established that ncRNAs in fact dominate the transcriptome and are implicated in a plethora of regulatory roles.
- Massive differences in coverage and biases in annotation between fairly closely related organisms.
- The list of non-protein coding RNAs seems still largely incomplete.

ENSEMBL 45 annotation of human and teleost fish genomes.

2. Prediction of Structured RNAs

- Applied genomes:
  - Takifugu rubripes, fugu, Ti
  - Tribolodon magnivelaris, Pacific silverside, Ts
  - Gasterosteus aculeatus, stickleback, Ga
  - Ornithorhynchus anatinus, platypus, Op

- Revealing undescribed, clade-specific, non-coding RNAs.

- Fish Specific Genome Duplication

- ncRNA candidates of unknown function.

- miRBase [3] is used to build genome-wide alignments of non-coding regions of the five teleosts.

- ** ncRNA candidates of unknown function.

- ** miRBase [3] is used to build genome-wide alignments of non-coding regions of the five teleosts.

- ** miRBase [3] is used to build genome-wide alignments of non-coding regions of the five teleosts.

3. Sensitivity

Sensitivity of ** miRBase on the annotated gene sets. Estimated per teleost.

Comparison: ** vs. ** vs. **.

4. Novel microRNAs

- Altuvia et al. [3]: miRNA precursors occur in close vicinity (<1000 nt).
- 1 Kb cluster
- 10 Kb cluster

5. Genomic Clusters

Distribution of structure distances between pairs of adjacent high confidence (** **) ** hits with a maximum distance of 10k and 10k, resp.

- All genomic clusters of ** predictions
- Restricted to the clusters containing at least one unannotated signal
- Completely unannotated clusters only

6. Paralogs

Density plot of the distribution of all pairs' LocARNA distances of putatively duplicated pairs with recognizable sequence similarity.

A (red curve), B: duplicated pairs, A (black curve), C: background.

7. Distribution of structure distances

- Unbiased survey for evolutionary conserved structured ncRNAs in teleost fish genomes
- Evidence for several thousand structured RNA motifs
- Only a small fraction can be annotated
- Strong evidence for the existence of previously undescribed structurally defined ncRNA families from structure-based clustering
- Very few ncRNAs retain recognizable duplicates
- Large scale duplication events do not lead to a corresponding increase in the ncRNA repertoire (at least as far as RNAs are concerned that depend on a well-defined structure.)
- One immediate implication is that comparative approaches within the same genome, i.e., comparisons between paralogous regions, will have very limited sensitivity at least for ncRNA discovery.

8. Conclusions

Acknowledgements

- For contributing:
  - Julian Jiric, Jorg Hackermüller, Kristin Reiche, Qiang Li, Peter F. Stadler
- For funding:
  - German DFG Bioinformatics Initiative IB2/1-2
  - 6th Framework Programme of the European Union (SYNLEET)
  - Vereinigung von Förderern und Freunden der Universität Leipzig e.V.
- Prof. Bulent Hsu, Fudan University

References