RNAcentral: Progress in the Development of the Non-coding RNA Sequence Database

Other potential titles:

- RNAcentral: New Developments in the Non-coding RNA Sequence Database
- RNAcentral: Three Years of ncRNA Data Integration

Anton I. Petrov[1], Blake A. Sweeney[1], Boris Burkov[1*], Simon Kay[1], Rob Finn[1], Alex Bateman[1], and The RNAcentral Consortium

European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD
*To whom correspondence should be addressed: burkov@ebi.ac.uk

BACKGROUND

More than fifty major specialised resources exist in the area of non-coding RNA (ncRNA) research.

RNAcentral (http://rnacentral.org) provides a unified interface that aggregates information from over twenty of them, including:

- larger databases (e.g. RefSeq, ENA, Ensembl!)
- databases with specific content type (e.g. Rfam, PDBe)
- RNA type-specific databases (e.g. miRBase, GtRNAdb, snoPY)
- organism-specific databases (e.g. HGNC, FlyBase, PomBase).

RESULTS

Launched in 2014 [2], RNAcentral currently contains over 10 million ncRNA sequences from more than 20 RNA databases and Is updated several times a year.

Recent updates include ncRNA data from:

- HGNC
- Ensembl
- FlyBase
- Rfam (most of sequences in RNAcentral are annotated with Rfam families, while the rest are candidates for producing new Rfam families)

There are three main ways of browsing the data through the RNAcentral website.

- The text search makes it easy to explore all ncRNA sequences, compare data across different resources, and discover what is known about each ncRNA.
- Using the sequence similarity search one can search data from multiple RNA databases starting from a sequence.
- Finally, one can explore ncRNAs in select species by genomic location using an integrated genome browser.

The data can also be accessed programmatically using an API or downloaded for local analysis from the FTP archive. We are looking for ways to improve the service, and we invite users to submit feedback at http://rnacentral.org/contact.

CONCLUSIONS

Recent development of RNAcentral includes importing new data and developing new website functionality. Our immediate priorities include the incorporation of functional annotations of non-coding RNAs, such as intermolecular interactions, nucleotide modifications, and high-quality secondary structures. The ultimate goal of RNAcentral is to include curated information about all non-coding RNAs as UniProt does for proteins.

REFERENCES

1. The RNAcentral Consortium. (2017) RNAcentral: a comprehensive database of non-coding RNA sequences.
   Nucleic Acids Res.

2. The RNAcentral Consortium. (2015) RNAcentral: an international database of ncRNA sequences. Nucleic Acids Res.
3. Nawrocki EP, Burge SW, Bateman A, et al (2015) Rfam 12.0: updates to the RNA families database.
Nucleic Acids Res.