Rfam: Quick tour

Sarah Burge [1]

- DNA & RNA
- Beginner
- 0.5 hour

This quick tour provides a brief introduction to Rfam, the EBI's database of RNA families.

Learning objectives:

- Basic understanding of what Rfam provides.
- Searching and discovering RNA-related information with Rfam.
- Know where to find out more about Rfam.

What is Rfam?

Rfam [2] is a database of non-coding RNA [3]. We classify non-coding RNAs into families, using multiple sequence alignments, consensus secondary structure [4] and covariance models [5]. Based on known members of a given RNA family, we create mathematical models that are used to computationally search sequence databases to identify new members of the family.

Each family in Rfam is assigned to one of three categories:

- non-coding RNA genes;
- structured cis-regulatory elements [6];
- self-splicing introns.

These are further subdivided into more specialised categories. We do not include piRNAs or short RNAs.

Why do we need Rfam?

Rfam was established to make it easier to identify homologues of known non-coding RNA. We also provide comprehensive text and functional annotation [7] for these sequences. Computational analysis of non-coding RNAs is time consuming when compared to doing similar searches on protein. We provide access to RNA search tools and results which require a lot of computational time and might not otherwise be accessible to scientists.
What is an Rfam family?

Each family starts with a group of non-coding RNA [3] sequences which are known to be true members of a family. These sequences are usually taken from the literature, or from direct submissions from our users. Collectively, a multiple sequence alignment of these sequences is referred to as the seed alignment. The seed sequences and alignments are manually curated by us.

Unlike proteins, RNAs belonging to the same family can have relatively weak sequence conservation, but can exhibit a high degree of secondary structure [4] conservation (See Figure 1). In order to look for similarities both in sequence and structure for RNAs, we use a type of model called a covariance model. Covariance models [5] (abbreviated to CM) may be thought of as more complicated hidden Markov models.

Figure 1 Example seed sequence alignment, showing the secondary structure regions (coloured regions).

We build a covariance model from the seed alignment. We then use this model to search our sequence database for sequences to include in the family. Those sequences which we find using the CM belong to the full set of family members.

We provide sequence match lists for the full members of the family, images of the secondary structure, as well as gene and sequence ontology [8] terms for each family. Text annotation [7] describing the family’s function is provided through Wikipedia.

What can I do with Rfam?

With Rfam [9] you can:

- Search our RNAs by keyword or family type;
- search your DNA or RNA sequence against our family models;
- browse our families and clans;
- retrieve text annotation [7] about a given family;
- explore the family's seed alignment and secondary structure [4];
- Retrieve match lists for the full membership;
- View family according to taxonomy.

Searching and visualising data in Rfam
Text search in Rfam

From the Rfam homepage [10] you can search for an Rfam accession [11] or identifier [12] directly. Rfam accessions are unique identifiers in the form RF0XXX, and identifiers are short text strings such as 5S_rRNA or mir-127. Use the "Jump to" box in the middle of the homepage to search on either of these identifiers. The "Jump to" box can also be used to search for Rfam clans.

You can also perform a keyword search using the box at the top right of every page.
Search a nucleic acid sequence against Rfam

If you have a nucleic acid sequence, you can search it against all the families in Rfam [9], to see if it matches any of them. Click on the “Search” link at the top of every page, and paste your sequence into the text box. Hit “Submit” to start your search running.
Rfam: Quick tour
Published on EMBL-EBI Train online (https://www.ebi.ac.uk/training/online)

Figure 3 Search box for Rfam sequence search.

Once your search has finished running, the results are automatically displayed (see Figure 4 below).

Figure 4 Example search results page.

Click on the links in the "Id" or "Accession [11]" field to access the family for a given result.

Using Rfam search results

Clicking on a family link (either from a sequence or text/keyword search result page) takes you to an Rfam [9] family home page. From here, the sidebar contains links to the data we provide - secondary structure [4], seed alignments for download, taxonomy browser and more. You will also see our free text annotation [7] provided by Wikipedia.
Figure 5 An Rfam family home page.

Getting data from Rfam

Bulk download of data

You can download the entire Rfam [9] MySQL [13] database dump from our FTP site [14]. Here you will also find other bulk downloads, such as the CM library, all Rfam seed sequences and mapping files.

Programmatic access to Rfam

We provide a RESTful interface for programmatic access to Rfam. More information may be found on our help pages [15].

Submitting data to Rfam

If you wish to submit either a new family or a new alignment for an existing family please email us on rfam [9]-help [at] ebi.ac.uk.

If you wish to submit a new annotation [7] for an existing family in Rfam you can edit the Wikipedia article linked to from the family. We provide a link from the family page for this purpose.
Figure 6  Example family page showing the link to 'Edit Wikipedia article'.

Get help and support on Rfam

Find out more...
You can check out our blog posts [16] or follow us on Twitter [17].

Support

- Rfam [9]-help [at] ebi.ac.uk (Email) the Rfam support team with your queries.
- Our help pages [18] contain further information about Rfam.

References

If you use Rfam in your work, please cite our most recent paper [19].

Collaborators

Rfam is based on the Infernal suite of software, produced by Eric Nawroki and Sean Eddy at HHMI Janelia Farm [20].

Contributors

This page populates automatically: please send me a short biography.
Sarah Burge [1]
EMBL-EBI
Visitor - Finn team: Protein families

Ruth Eberhardt [21]
EMBL-EBI
Visitor - Finn team: Protein families

Source URL: https://www.ebi.ac.uk/training/online/course/rfam-quick-tour

Links
[1] https://www.ebi.ac.uk/training/online/trainers/swb
[3] https://www.ebi.ac.uk/training/online/glossary/non-coding-rna
[4] https://www.ebi.ac.uk/training/online/glossary/secondary-structure
[5] https://www.ebi.ac.uk/training/online/glossary/covariance-models
[6] https://www.ebi.ac.uk/training/online/glossary/cis-regulatory-elements
[7] https://www.ebi.ac.uk/training/online/glossary/annotation
[8] https://www.ebi.ac.uk/training/online/glossary/ontology
[9] https://www.ebi.ac.uk/training/online/glossary/rfam
[11] https://www.ebi.ac.uk/training/online/glossary/accession
[12] https://www.ebi.ac.uk/training/online/glossary/identifier
[13] https://www.ebi.ac.uk/training/online/glossary/mysql
[17] https://twitter.com/Xfam_EBI
[18] http://rfam.xfam.org/help#tabview=tab0
[21] https://www.ebi.ac.uk/training/online/trainers/ruthe