UC Santa Cruz

UC Santa Cruz Previously Published Works

Title

RNAcentral: a comprehensive database of non-coding RNA sequences.

Permalink

https://escholarship.org/uc/item/22g1v0g8

Journal

Nucleic Acids Research, 45(D1)

Authors

Petrov, Anton

Kay, Simon

Kalvari, Ioanna

et al.

Publication Date

2017-01-04

DOI

10.1093/nar/gkw1008

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Nucleic Acids Research, 2016 **1**
doi: 10.1093/nar/gkw1008

RNAcentral: a comprehensive database of non-coding RNA sequences

The RNAcentral Consortium[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,*]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, [2]Faculty of Biology, Medicine and Health, University of Manchester, Oxford Road, Manchester M13 9PT, UK, [3]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1HH, UK, [4]Department of Biochemistry, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229-3900, USA, [5]Sandia National Laboratories, Livermore, CA 94551, USA, [6]Department of Biomolecular Engineering, University of California Santa Cruz, CA 95064, USA, [7]Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX 78712, USA, [8]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland, [9]Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, Umultowska 89, 61-614 Poznan, Poland, [10]Frontier Science Research Center, University of Miyazaki, Japan, [11]Michigan State University, East Lansing, MI 48824-1325, USA, [12]Department of Computational Biology, Adam Mickiewicz University in Poznan, Poland, [13]Cambridge Systems Biology Centre & Department of Biochemistry, University of Cambridge, Sanger Building, 80 Tennis Court Road, Cambridge CB2 1GA, UK, [14]The Arabidopsis Information Resource and Phoenix Bioinformatics, 643 Bair Island Rd. Suite 403, Redwood City, CA 94063, USA, [15]Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China, [16]Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China, [17]Data Science, National Center for Protein Science, Beijing, China, [18]DIANA-Lab, Department of Electrical & Computer Engineering, University of Thessaly, 382 21 Volos, Greece, [19]Hellenic Pasteur Institute, 127 Vasilissis Sofias Avenue, 11521 Athens, Greece, [20]BIG Data Center and CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [21]University of Strasbourg, 15 rue R. Descartes, 67084 Strasbourg, France, [22]Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, D-04107 Leipzig, Germany, [23]Department of Pediatrics, University of California San Diego, La Jolla, CA, USA, [24]dictyBase, Northwestern University, Chicago, IL, USA, [25]Department of Genetics, Stanford University, Stanford, CA 94305, USA, [26]Center for Medical Genetics, Ghent University and Cancer Research Institute Ghent, Ghent University, Ghent, Belgium, [27]Department of Animal Sciences, Auburn University, Auburn, AL 36849, USA, [28]Garvan Institute of Medical Research, Sydney, New South Wales 2010, Australia and [29]MRC Functional Genomics Unit, Department of Physiology, Anatomy, and Genetics, University of Oxford, Oxford OX1 3PT, UK

Received September 19, 2016; Revised October 13, 2016; Editorial Decision October 14, 2016; Accepted October 18, 2016

ABSTRACT

RNAcentral is a database of non-coding RNA (ncRNA) sequences that aggregates data from specialised ncRNA resources and provides a single entry point for accessing ncRNA sequences of all ncRNA types from all organisms. Since its launch in 2014, RNAcentral has integrated twelve new resources, taking the total number of collaborating database to 22, and began importing new types of data, such as modified nucleotides from MODOMICS and PDB. We created new species-specific identifiers that refer to unique RNA sequences within a context of single species. The website has been subject to continuous improvements focusing on text and sequence similarity searches as well as genome browsing functionality. All RNAcentral data is provided for free and is available for browsing, bulk downloads, and programmatic access at http://rnacentral.org/.

INTRODUCTION

Non-coding RNAs (ncRNAs) are a critical component of cellular machinery of all organisms. For example, the ribosomal RNA has been shown to be a ribozyme responsible for peptide bond synthesis (1), and the activity of the eukaryotic spliceosome is mediated by ncRNAs (2). Apart

*To whom correspondence should be addressed. Tel: +44 1223 492550; Fax: +44 1223 494468; Email: apetrov@ebi.ac.uk
Present address: Magdalena A. Machnicka, Faculty of Mathematics, Informatics and Mechanics (MIM), University of Warsaw, Banacha 2, 02-097 Warsaw, Poland.

© The Author(s) 2016. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

from being the main player in those central processes, ncRNAs provide additional layers of subtle regulation of gene expression. MicroRNAs have been shown to regulate the expression of the majority of mRNAs in animals and plants (3), and the range of regulatory roles of lncRNAs, including by genomic scaffolding and chromatin remodelling and modification (4), is becoming clearer. There is an intense scientific interest in ncRNAs resulting in a large number of ncRNA databases, but until recently searching and comparing them was challenging, and there was no uniform way to access or reference ncRNA sequences. To this end, we developed RNAcentral, a database of ncRNA sequences that serves as a single entry point to the data from a large collection of collaborating ncRNA resources that cover ncRNA sequences of all types and from all organisms. First conceived in 2011 (5), RNAcentral was made public in 2014 (6). This paper gives an update on the status of the database and related activities.

DATA OVERVIEW

New Expert Databases

RNAcentral aggregates ncRNA sequence data from an international consortium of RNA resources that we call Expert Databases. In the past two years, 12 additional Expert Databases have been integrated into RNAcentral (see Table 1). Among the newly imported resources were two major ribosomal RNA databases, SILVA (7) and Greengenes (8), which complement rRNAs from ENA and Rfam, as well as a high quality subset of rRNA sequences from RDP (9). Ribosomal RNAs represent the majority of sequences in RNAcentral due to their use in environmental sampling.

Sequences from five Model Organism Databases (DictyBase (10), PomBase (11), SGD (12), TAIR (13) and WormBase (14)) have been imported into RNAcentral. Long noncoding RNA (lncRNA) coverage was extended by the addition of NONCODE (15) and LNCipedia (16) datasets. The inclusion of PDB (17) as an Expert Database helps to map between the worlds of ncRNA sequence and structure. Small nucleolar RNAs (snoRNAs) play an important role in guiding the modification of other ncRNA and mRNAs (18), and snoPY (19) provides to RNAcentral a dataset of snoRNAs found in human, fly, worm, yeast, and thale cress. An up-to-date list of all RNAcentral Expert Databases is available at http://rnacentral.org/expert-databases.

Database growth

RNAcentral currently holds 10.2 million distinct ncRNA sequences (an increase of 2.1 million since release 1) with about 28 million cross-references (up 11 million since release 1) to 22 Expert Databases (Figure 1). The sequences come from over 720 000 organisms from all domains of life, with half of all sequences deriving from bacteria and ∼40% from eukaryotes.

Species-specific identifiers

Since its inception, RNAcentral has provided unique identifiers for each distinct RNA sequence. For example, URS00004C905 is the identifier for the sequence UU UGGUCCCCUUCAACCAGCUG, which is the miRNA miR-133a-3p. The exact same sequence is observed in more than 10 species. These unique and stable identifiers are useful for a number of tasks, such as unambiguously referring to an RNA sequence, reducing redundancy in sequence datasets, and keeping track of cross-references. However, it is essential to be able to refer uniquely to a ncRNA sequence annotation in a particular species. In order to address this issue, we have introduced species-specific RNA sequence identifiers. The new identifiers are composed of a unique RNA sequence identifier and a NCBI taxonomic identifier separated by underscore. For example, URS00004C9052_9606 is the human copy of the miRNA hsa-miR-133a-3p (*9606* being the taxonomic identifier for *Homo sapiens*) and URS00004C9052_10090 corresponds to the mmu-miR-133a-3p sequence from mouse. RNAcentral text searches now return species-specific entries by default.

One of the use cases for the species-specific RNA sequence identifiers is curation of literature references to assign ontology terms to an RNA sequence from a specific organism, for example to annotate the molecule's function. The biocuration community has already begun using RNAcentral identifiers for assigning Gene Ontology terms to human miRNAs (20). We are investigating assigning identifiers that differentiate between multiple occurrences of the same sequence in a genome.

Modified nucleotides

Modified nucleotide residues play important roles in functions of many ncRNAs. For example, modifications of ribosomal RNAs are essential for the assembly and stability of ribosomes (21), and tRNA modifications can influence protein gene expression (22). In order to begin capturing information about modified residues in RNA molecules and enable comparison of different datasets, we imported modified rRNA and tRNA sequences from MODOMICS (23), a database of RNA modification pathways, as well as all ncRNA modifications from Protein Data Bank. So far RNAcentral contains over 170 different chemical modifications found at over 8000 positions in about 600 unique sequences. Figure 2 shows a web interface that provides a unified view of the modifications from different databases. For each modified residue, cross-references to the MODOMICS and PDB databases are provided to enable easy access to more detailed information about each modification. In future releases we will continue importing information about modified nucleotides from MODOMICS and other resources as more data become available thanks to new developments in sequencing technology (24,25).

WEBSITE UPDATES

The RNAcentral website has been subject to continuous improvement based on user feedback and several interactive workshops held during annual RNAcentral consortium meetings. The homepage was redesigned to reflect the three main ways to access the data: **text search, sequence similarity search** and **genome browser**, each of which will be discussed further below.

Table 1. Expert Databases imported into RNAcentral since release 1

Database name	Description	URL
DictyBase	A model organism database for the social amoeba *Dictyostelium discoideum*	http://dictybase.org
Greengenes	A full-length 16S rRNA gene database that provides a curated taxonomy based on de novo tree inference	http://greengenes.secondgenome.com/downloads
LNCipedia	An integrated database of human lncRNAs	http://www.lncipedia.org
MODOMICS	A comprehensive database of RNA modifications	http://modomics.genesilico.pl
NONCODE	An integrated knowledge database dedicated to ncRNAs (excluding tRNAs and rRNAs)	http://www.noncode.org
PDB	A repository of information about the 3D structures of large biological molecules	http://www.wwpdb.org/
PomBase	A comprehensive database for the fission yeast *Schizosaccharomyces pombe*	http://www.pombase.org
SGD	An integrated database for the budding yeast	http://yeastgenome.org
SILVA	A resource for quality checked and aligned ribosomal RNA sequence data	http://www.arb-silva.de/
snoPY	A database of snoRNAs, snoRNA gene loci, and target RNAs as well as snoRNA orthologues	http://snoopy.med.miyazaki-u.ac.jp/
TAIR	A database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*	http://www.arabidopsis.org
WormBase	A resource for genomic and genetic data about nematodes with primary emphasis on *Caenorhabditis elegans*	http://www.wormbase.org

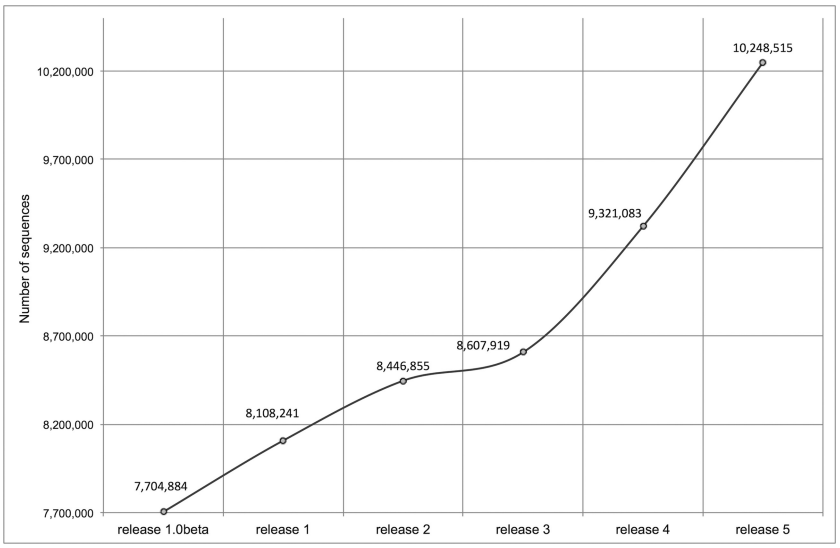

Figure 1. Growth in the number of unique RNA sequences since release 1. An up-to-date version of the chart is available at http://rnacentral.org/about-us.

Text search

RNAcentral text search provides a flexible way for exploring RNAcentral data using a faceted interface powered by EBI search (26). In the past two years, the search functionality was improved both in terms of user interface and searchable data. For example, publication metadata (such as paper titles, PubMed identifiers or author names) associated with ncRNA sequences can now be searched, which makes it possible to look up ncRNA sequences submitted to sequence archives when new papers are published. For example, a recent paper describes TRM10 (27), a mRNA-derived small RNA that acts as ribosome inhibitor. By searching RNAcentral with PubMed identifier *24685157* the sequence is easily found (see entry URS00007E15D1) and can be used for further analysis, such as sequence similarity search. Moreover, it is possible to compare sequences reported in different papers. For example, one can identify mitochondrial rRNA sequences shared by a Danish (28) and an Iranian (29) population by searching in RNAcentral with both publication titles. These search results can be exported in multiple formats, thereby facilitating more detailed investigation by the user.

Sequence similarity search

RNAcentral sequence search is the first online tool that enables sequence similarity searches against a comprehensive set of ncRNAs. The service is powered by the *nhmmer* software, which has a comparable speed to BLAST but is more sensitive (30). The web interface supports searching using an RNA or DNA sequence as a query and displays pairwise sequence alignments for each match. The results can be sorted by E-value, sequence identity and other criteria. If an exact match for a query sequence already exists in the database, its RNAcentral identifier is retrieved using the

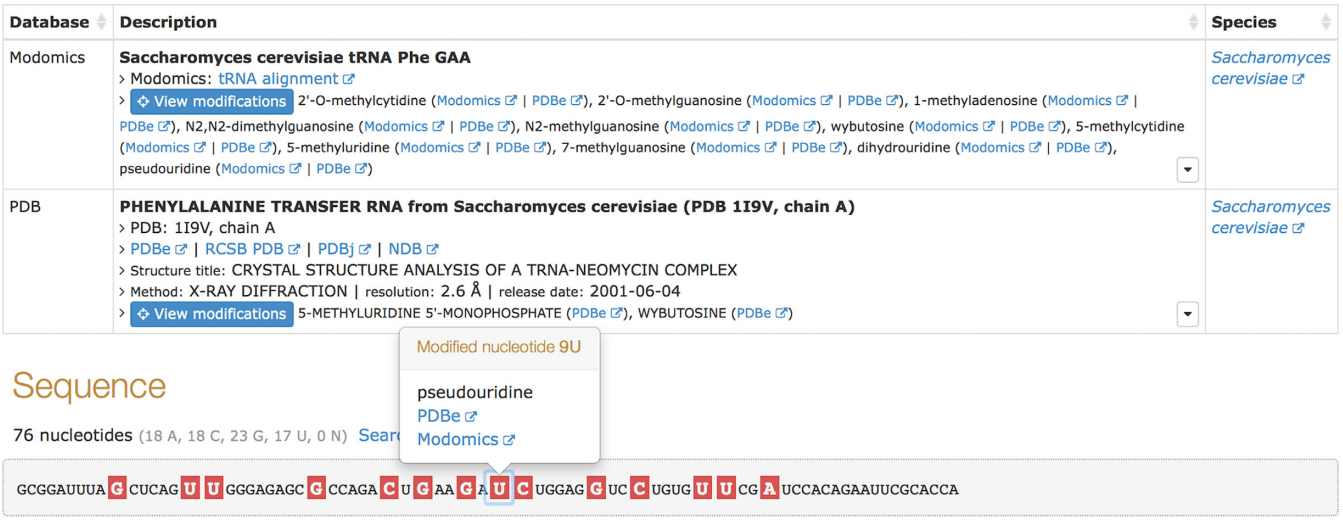

Figure 2. Web interface displaying modified nucleotides for a *Saccharomyces cerevisiae* tRNA(Phe) sequence (RNAcentral entry URS000011107D_4932).

RNAcentral API without having to wait for the full search results to become available.

Genome browser

Viewing sequences in their genomic context can provide important biological insights. For example, one can visualise snoRNAs found in introns of lncRNA GAS5 (31) using a built-in genome browser (see RNAcentral entry URS00008B3C85). In a recent update, we extended this functionality to enable browsing RNAcentral starting with a genomic location. The embedded genome browser, powered by Genoverse (http://genoverse.org), currently supports 13 key species, including human, mouse, fly, worm, and yeast (Figure 3). RNAcentral sequences are displayed alongside genes and transcripts from Ensembl (32) and Ensembl Genomes (33) with links to fully featured genome browsers, such as UCSC (34) and Ensembl. The genomic data are also available via a programmatic interface and downloadable files in BED/GFF formats.

RNACENTRAL USE CASES

Citations to RNAcentral are beginning to appear in the scientific literature, and currently fall into three main categories of use: (i) RNAcentral is used as a comprehensive source of ncRNA annotations and a reference for identification of novel ncRNAs in species like rainbow trout or cow (35,36). (ii) RNAcentral identifiers are used for literature curation, for example, human miRNAs are annotated with GO terms using RNAcentral species-specific identifiers to refer to RNA sequences (20). (iii) The RNAcentral API is used for sequence or identifier retrieval (37–39). For example, given an RNAcentral sequence identifier the Forna tool can predict and visualise its secondary structure. Over the past two years, the RNAcentral website has been accessed by over 33 000 unique visitors from 156 countries who performed over 100 000 text and 12 000 sequence similarity searches.

TRAINING AND OUTREACH

We continuously engage in outreach activities and provide user support by email and on GitHub. We have delivered over 20 presentations to date at scientific conferences and research institutes worldwide and organised a training event at the Wellcome Genome Campus. We also developed an online training course and recorded a live webinar (available on YouTube). All training materials can be accessed at http://rnacentral.org/training. We are open to suggestions from our user community by email, on GitHub and on Twitter. The contact information can be found at http://rnacentral.org/contact.

FUTURE DIRECTIONS

The main goal of RNAcentral is to provide a comprehensive set of ncRNA sequences, so integrating new Expert Databases and importing more data will remain a priority. For example, our goal is to integrate the remaining Model Organism Databases, such as FlyBase (40) and RGD (41), in order to provide uniform access to high-quality ncRNA sequence and annotations from key species. More than 20 participating ncRNA resources still need to be integrated and new Expert Databases are continually joining the consortium. We welcome relevant databases to contact us about membership.

In the second phase of development, we will begin to integrate new types of annotations that can provide insight into the functions of ncRNA sequences found in RNAcentral. We will work on importing secondary structure information from Comparative RNA Website (42), GtRNAdb (43) and Rfam (44) databases. Work is underway on integrating miRNA-mRNA interactions from TarBase (45) and miRNA–lncRNA interactions from LncBase (46) into RNAcentral. We will enrich existing annotations by importing ontology terms from external resources and assigning ontology terms automatically where possible. We also plan to use sequence alignment-based mapping to connect more RNAcentral sequences to reference genomes. RNAcentral

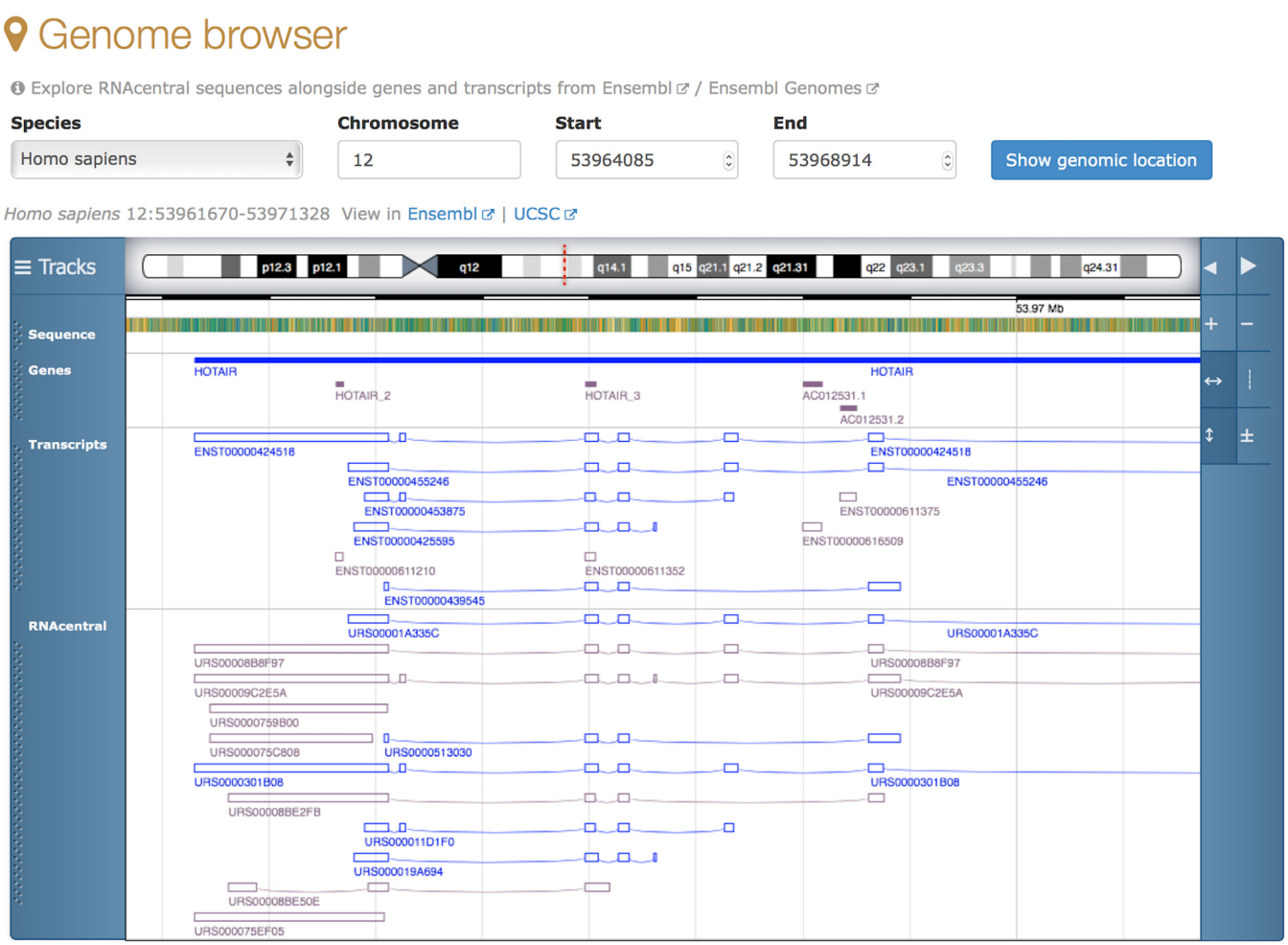

Figure 3. RNAcentral genome browser showing HOTAIR lncRNA in human chromosome 12.

is a young and fast-growing resource, but it has already proved useful for many applications, and its utility will be increased as more data are integrated and the associated services mature.

ACKNOWLEDGEMENTS

RNAcentral has been prepared by Anton I. Petrov, Simon J.E. Kay, Ioanna Kalvari, Kevin L. Howe, Kristian A. Gray, Elspeth A. Bruford, Paul J. Kersey, Guy Cochrane, Robert D. Finn, Alex Bateman at the European Bioinformatics Institute (EMBL-EBI); Ana Kozomara, Sam Griffiths-Jones (University of Manchester); Adam Frankish (Wellcome Trust Sanger Institute); Christian W. Zwieb (University of Texas), Britney Y. Lau, Kelly P. Williams (Sandia National Laboratories); Patricia P. Chan, Todd M. Lowe (University of California Santa Cruz); Jamie J. Cannone, Robin R. Gutell (University of Texas at Austin); Magdalena A. Machnicka, Janusz M. Bujnicki (International Institute of Molecular and Cell Biology and Adam Mickiewicz University); Maki Yoshihama, Naoya Kenmochi (University of Miyazaki); Benli Chai, James R. Cole (Michigan State University); Maciej Szymanski, Wojciech M. Karlowski (Adam Mickiewicz University); Valerie Wood (University of Cambridge); Eva Huala, Tanya Z. Berardini (The Arabidopsis Information Resource and Phoenix Bioinformatics); Yi Zhao, Runsheng Chen (Chinese Academy of Sciences); Weimin Zhu (Data Science, National Center for Protein Science); Maria D. Paraskevopoulou, Ioannis S. Vlachos, Artemis G Hatzigeorgiou (University of Thessaly and Hellenic Pasteur Institute); SILVA team (Jacobs University Bremen and Max Planck Institute for Marine Microbiology); Lina Ma, Zhang Zhang (Beijing Institute of Genomics, Chinese Academy of Sciences); Joern Puetz (University of Strasbourg); Peter F. Stadler (University of Leipzig); Daniel McDonald (University of California San Diego); Siddhartha Basu, Petra Fey (Northwestern University); Stacia R. Engel, J. Michael Cherry (Stanford University); Pieter-Jan Volders, Pieter Mestdagh (Ghent University and Cancer Research Institute Ghent); Jacek Wower (Auburn University); Michael Clark (University of Oxford and Garvan Institute of Medical Research); Xiu Cheng Quek, Marcel E. Dinger (Garvan Institute of Medical Research).

FUNDING

Biotechnology and Biological Sciences Research Council (BBSRC) [BB/J019232/1]. Funding for open access charge: Research Councils UK (RCUK).
Conflict of interest statement. Janusz M. Bujnicki is an Executive Editor of *Nucleic Acids Research*.

REFERENCES

1. Beringer,M. and Rodnina,M.V. (2007) The ribosomal peptidyl transferase. *Mol. Cell*, **26**, 311–321.
2. Hang,J., Wan,R., Yan,C. and Shi,Y. (2015) Structural basis of pre-mRNA splicing. *Science*, **349**, 1191–1198.
3. Axtell,M.J., Westholm,J.O. and Lai,E.C. (2011) Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biol.*, **12**, 221.
4. Tomita,S., Abdalla,M.O.A., Fujiwara,S., Yamamoto,T., Iwase,H., Nakao,M. and Saitoh,N. (2016) Roles of long noncoding RNAs in chromosome domains. *Wiley Interdiscip. Rev. RNA*, doi:10.1002/wrna.1384.
5. Bateman,A., Agrawal,S., Birney,E., Bruford,E.A., Bujnicki,J.M., Cochrane,G., Cole,J.R., Dinger,M.E., Enright,A.J., Gardner,P.P. *et al.* (2011) RNAcentral: A vision for an international database of RNA sequences. *RNA*, **17**, 1941–1946.
6. Consortium, RNAcentral (2015) RNAcentral: an international database of ncRNA sequences. *Nucleic Acids Res.*, **43**, D123–D129.
7. Quast,C., Pruesse,E., Yilmaz,P., Gerken,J., Schweer,T., Yarza,P., Peplies,J. and Glöckner,F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.*, **41**, D590–D596.
8. McDonald,D., Price,M.N., Goodrich,J., Nawrocki,E.P., DeSantis,T.Z., Probst,A., Andersen,G.L., Knight,R. and Hugenholtz,P. (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.*, **6**, 610–618.
9. Cole,J.R., Wang,Q., Fish,J.A., Chai,B., McGarrell,D.M., Sun,Y., Brown,C.T., Porras-Alfaro,A., Kuske,C.R. and Tiedje,J.M. (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.*, **42**, D633–D642.
10. Basu,S., Fey,P., Pandit,Y., Dodson,R., Kibbe,W.A. and Chisholm,R.L. (2013) DictyBase 2013: integrating multiple Dictyostelid species. *Nucleic Acids Res.*, **41**, D676–D683.
11. McDowall,M.D., Harris,M.A., Lock,A., Rutherford,K., Staines,D.M., Bähler,J., Kersey,P.J., Oliver,S.G. and Wood,V. (2015) PomBase 2015: updates to the fission yeast database. *Nucleic Acids Res.*, **43**, D656–D661.
12. Cherry,J.M., Hong,E.L., Amundsen,C., Balakrishnan,R., Binkley,G., Chan,E.T., Christie,K.R., Costanzo,M.C., Dwight,S.S., Engel,S.R. *et al.* (2012) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.*, **40**, D700–D705.
13. Berardini,T.Z., Reiser,L., Li,D., Mezheritsky,Y., Muller,R., Strait,E. and Huala,E. (2015) The Arabidopsis information resource: Making and mining the 'gold standard' annotated reference plant genome. *Genesis*, **53**, 474–485.
14. Yook,K., Harris,T.W., Bieri,T., Cabunoc,A., Chan,J., Chen,W.J., Davis,P., de la Cruz,N., Duong,A., Fang,R. *et al.* (2012) WormBase 2012: more genomes, more data, new website. *Nucleic Acids Res.*, **40**, D735–D741.
15. Zhao,Y., Li,H., Fang,S., Kang,Y., Wu,W., Hao,Y., Li,Z., Bu,D., Sun,N., Zhang,M.Q. *et al.* (2016) NONCODE 2016: an informative and valuable data source of long non-coding RNAs. *Nucleic Acids Res.*, **44**, D203–D208.
16. Volders,P.-J., Verheggen,K., Menschaert,G., Vandepoele,K., Martens,L., Vandesompele,J. and Mestdagh,P. (2015) An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res.*, **43**, D174–D180.
17. Velankar,S., van Ginkel,G., Alhroub,Y., Battle,G.M., Berrisford,J.M., Conroy,M.J., Dana,J.M., Gore,S.P., Gutmanas,A., Haslam,P. *et al.* (2016) PDBe: improved accessibility of macromolecular structure data from PDB and EMDB. *Nucleic Acids Res.*, **44**, D385–D395.
18. Dupuis-Sandoval,F., Poirier,M. and Scott,M.S. (2015) The emerging landscape of small nucleolar RNAs in cell biology. *Wiley Interdiscip. Rev. RNA*, **6**, 381–397.
19. Yoshihama,M., Nakao,A. and Kenmochi,N. (2013) snOPY: a small nucleolar RNA orthological gene database. *BMC Res. Notes*, **6**, 426.
20. Huntley,R.P., Sitnikov,D., Orlic-Milacic,M., Balakrishnan,R., D'Eustachio,P., Gillespie,M.E., Howe,D., Kalea,A.Z., Maegdefessel,L., Osumi-Sutherland,D. *et al.* (2016) Guidelines for the functional annotation of microRNAs using the Gene Ontology. *RNA*, **22**, 667–676.
21. Osterman,I.A., Sergiev,P.V., Tsvetkov,P.O., Makarov,A.A., Bogdanov,A.A. and Dontsova,O.A. (2011) Methylated 23S rRNA nucleotide m2G1835 of Escherichia coli ribosome facilitates subunit association. *Biochimie*, **93**, 725–729.
22. Duechler,M., Leszczyńska,G., Sochacka,E. and Nawrot,B. (2016) Nucleoside modifications in the regulation of gene expression: focus on tRNA. *Cell. Mol. Life Sci.*, **73**, 3075–3095.
23. Machnicka,M.A., Milanowska,K., Osman Oglou,O., Purta,E., Kurkowska,M., Olchowik,A., Januszewski,W., Kalinowski,S., Dunin-Horkawicz,S., Rother,K.M. *et al.* (2013) MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res.*, **41**, D262–D267.
24. Cozen,A.E., Quartley,E., Holmes,A.D., Hrabeta-Robinson,E., Phizicky,E.M. and Lowe,T.M. (2015) ARM-seq: AlkB-facilitated RNA methylation sequencing reveals a complex landscape of modified tRNA fragments. *Nat. Methods*, **12**, 879–884.
25. Krogh,N., Jansson,M.D., Häfner,S.J., Tehler,D., Birkedal,U., Christensen-Dalsgaard,M., Lund,A.H. and Nielsen,H. (2016) Profiling of 2'-O-Me in human rRNA reveals a subset of fractionally modified positions and provides evidence for ribosome heterogeneity. *Nucleic Acids Res.*, doi:10.1093/nar/gkw482.
26. Squizzato,S., Park,Y.M., Buso,N., Gur,T., Cowley,A., Li,W., Uludag,M., Pundir,S., Cham,J.A., McWilliam,H. *et al.* (2015) The EBI Search engine: providing search and retrieval functionality for biological data from EMBL-EBI. *Nucleic Acids Res.*, **43**, W585–W588.
27. Pircher,A., Bakowska-Zywicka,K., Schneider,L., Zywicki,M. and Polacek,N. (2014) An mRNA-derived noncoding RNA targets and regulates the ribosome. *Mol. Cell*, **54**, 147–155.
28. Li,S., Besenbacher,S., Li,Y., Kristiansen,K., Grarup,N., Albrechtsen,A., Sparsø,T., Korneliussen,T., Hansen,T., Wang,J. *et al.* (2014) Variation and association to diabetes in 2000 full mtDNA sequences mined from an exome study in a Danish population. Eur. *J. Hum. Genet.*, **22**, 1040–1045.
29. Derenko,M., Malyarchuk,B., Bahmanimehr,A., Denisova,G., Perkova,M., Farjadian,S. and Yepiskoposyan,L. (2013) Complete mitochondrial DNA diversity in Iranians. *PLoS One*, **8**, e80673.
30. Wheeler,T.J. and Eddy,S.R. (2013) nhmmer: DNA homology search with profile HMMs. *Bioinformatics*, **29**, 2487–2489.
31. Smith,C.M. and Steitz,J.A. (1998) Classification of gas5 as a multi-small-nucleolar-RNA (snoRNA) host gene and a member of the 5'-terminal oligopyrimidine gene family reveals common features of snoRNA host genes. *Mol. Cell. Biol.*, **18**, 6897–6909.
32. Yates,A., Akanni,W., Amode,M.R., Barrell,D., Billis,K., Carvalho-Silva,D., Cummins,C., Clapham,P., Fitzgerald,S., Gil,L. *et al.* (2016) Ensembl 2016. *Nucleic Acids Res.*, **44**, D710–D716.
33. Kersey,P.J., Allen,J.E., Armean,I., Boddu,S., Bolt,B.J., Carvalho-Silva,D., Christensen,M., Davis,P., Falin,L.J., Grabmueller,C. *et al.* (2016) Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res.*, **44**, D574–D580.
34. Speir,M.L., Zweig,A.S., Rosenbloom,K.R., Raney,B.J., Paten,B., Nejad,P., Lee,B.T., Learned,K., Karolchik,D., Hinrichs,A.S. *et al.* (2016) The UCSC Genome Browser database: 2016 update. *Nucleic Acids Res.*, **44**, D717–D725.
35. Al-Tobasei,R., Paneru,B. and Salem,M. (2016) Genome-wide discovery of long non-coding RNAs in rainbow trout. *PLoS One*, **11**, e0148940.
36. Durkin,K., Rosewick,N., Artesi,M., Hahaut,V., Griebel,P., Arsic,N., Burny,A., Georges,M. and Van den Broeke,A. (2016) Characterization of novel Bovine Leukemia Virus (BLV) antisense transcripts by deep sequencing reveals constitutive expression in tumors and transcriptional interaction with viral microRNAs. *Retrovirology*, **13**, 33.

37. Kerpedjiev,P., Hammer,S. and Hofacker,I.L. (2015) Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics*, **31**, 3377–3379.
38. Jossinet,F., Ludwig,T.E. and Westhof,E. (2010) Assemble: an interactive graphical tool to analyze and build RNA architectures at the 2D and 3D levels. *Bioinformatics*, **26**, 2057–2059.
39. Eggenhofer,F., Hofacker,I.L. and Höner Zu Siederdissen,C. (2016) RNAlien - unsupervised RNA family model construction. *Nucleic Acids Res.*, doi:10.1093/nar/gkw558.
40. Attrill,H., Falls,K., Goodman,J.L., Millburn,G.H., Antonazzo,G., Rey,A.J., Marygold,S.J. and FlyBase Consortium (2016) FlyBase: establishing a Gene Group resource for Drosophila melanogaster. *Nucleic Acids Res.*, **44**, D786–D792.
41. Shimoyama,M., De Pons,J., Hayman,G.T., Laulederkind,S.J.F., Liu,W., Nigam,R., Petri,V., Smith,J.R., Tutaj,M., Wang,S.-J. *et al.* (2015) The Rat Genome Database 2015: genomic, phenotypic and environmental variations and disease. *Nucleic Acids Res.*, **43**, D743–D750.
42. Cannone,J.J., Subramanian,S., Schnare,M.N., Collett,J.R., D'Souza,L.M., Du,Y., Feng,B., Lin,N., Madabusi,L.V., Müller,K.M. *et al.* (2002) The comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics*, **3**, 1–31.
43. Chan,P.P. and Lowe,T.M. (2016) GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.*, **44**, D184–D189.
44. Nawrocki,E.P., Burge,S.W., Bateman,A., Daub,J., Eberhardt,R.Y., Eddy,S.R., Floden,E.W., Gardner,P.P., Jones,T.A., Tate,J. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.
45. Vlachos,I.S., Paraskevopoulou,M.D., Karagkouni,D., Georgakilas,G., Vergoulis,T., Kanellos,I., Anastasopoulos,I.-L., Maniou,S., Karathanou,K., Kalfakakou,D. *et al.* (2015) DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Res.*, **43**, D153–D159.
46. Paraskevopoulou,M.D., Vlachos,I.S., Karagkouni,D., Georgakilas,G., Kanellos,I., Vergoulis,T., Zagganas,K., Tsanakas,P., Floros,E., Dalamagas,T. *et al.* (2016) DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.*, **44**, D231–D238.