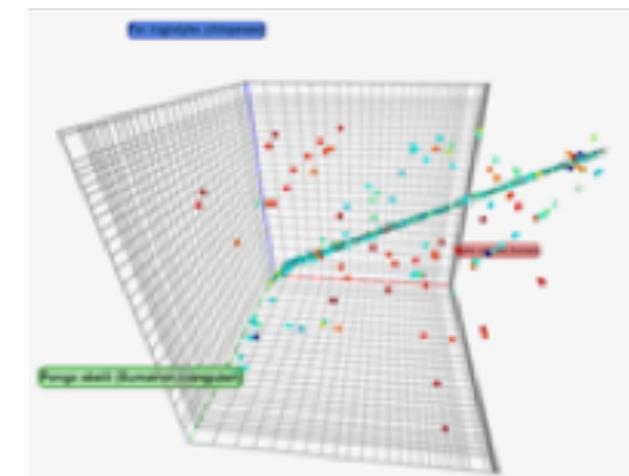
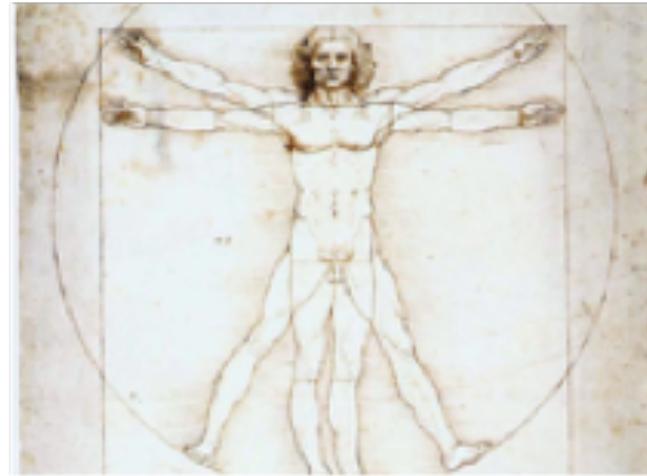
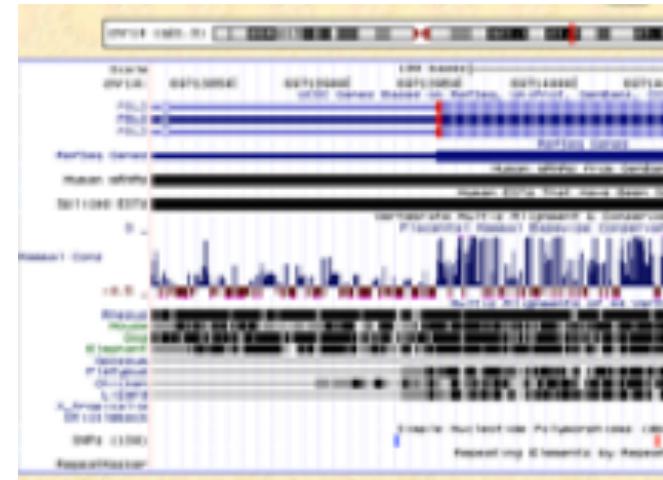


Computational Genomics

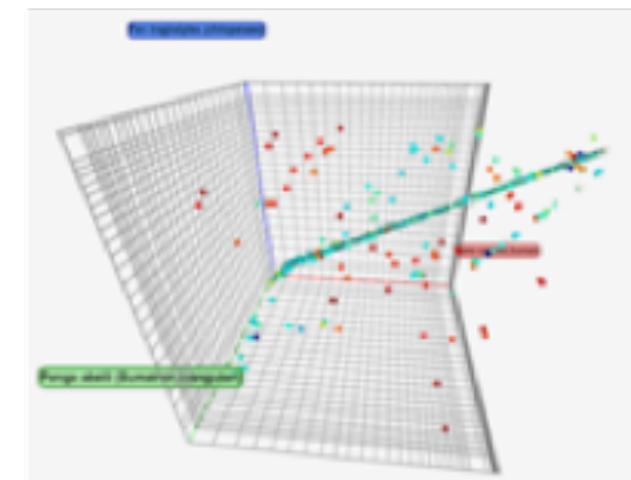
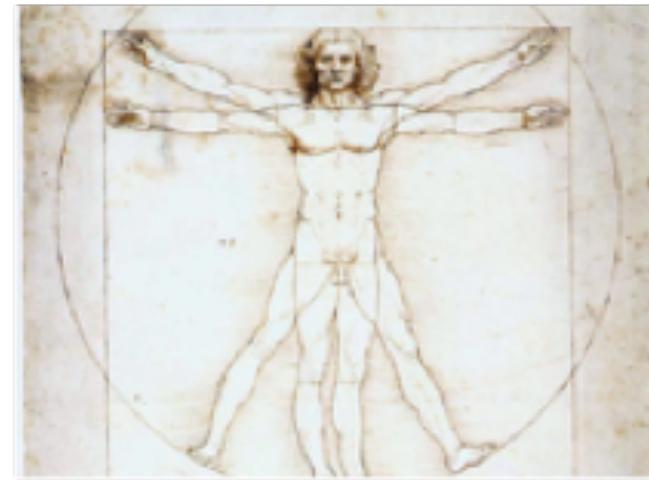
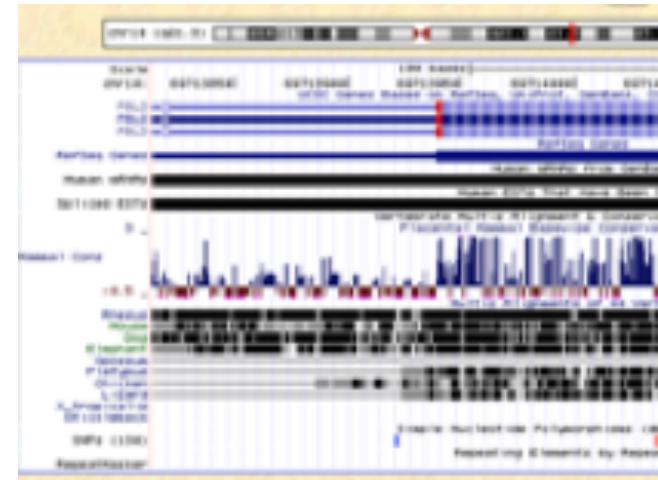
Introduction To Databases Finding: Genes, and Genomes Proteins and Proteomes



Computational Genomics

Introduction To Databases

DNA Databases



DNA Databases and Genome Browsers

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts
into NCBI databases



Download

Transfer NCBI data to your
computer



Learn

Find help documents, attend a
class or watch a tutorial



Develop

Use NCBI APIs and code
libraries to build applications



Analyze

Identify an NCBI tool for your
data analysis task



Research

Explore NCBI research and
collaborative projects



Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

Try out the new gene table from NCBI Datasets!

24 Jan 2022

Announcing a new feature in NCBI

Venturing beyond the genes: New RefSeq Functional Elements publication!

19 Jan 2022

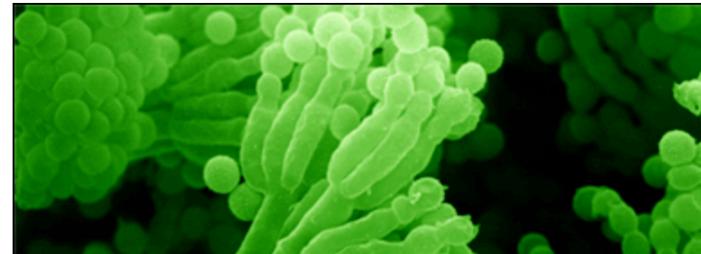
If you're curious about genome

RefSeq Release 210 is available

11 Jan 2022

RefSeq Release 210 is now available online, from the FTP site and through

DNA Databases and Genome Browsers



Bookshelf

Bookshelf provides free online access to books and documents in life science and healthcare. Search, read, and discover.

Using Bookshelf

[Quick Start Guide](#)

[FAQ](#)

[Tutorials](#)

[Copyright and Permissions](#)

[Follow @ncbibooks](#)

Read

[Browse Titles](#)

[New Releases](#)

[PubReader](#)

Participate

[Authors and Publishers](#)

[How to Apply](#)

[Participation Agreement](#)

[File Submission Specifications](#)

DNA Databases and Genome Browsers



Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

Using Genome

[Help](#)

[Browse by Organism](#) UPDATED

[Download / FTP](#)

[Download FAQ](#)

[Submit a genome](#)

Custom resources

[Human Genome](#)

[Microbes](#)

[Organelles](#)

[Viruses](#)

[Prokaryotic reference genomes](#)

Other Resources

[Assembly](#)

[BioProject](#)

[BioSample](#)

[Genome Data Viewer](#)

[NCBI Datasets](#) NEW

Genome Tools

[BLAST the Human Genome](#)

[Microbial Nucleotide BLAST](#)

Genome Annotation and Analysis

[Eukaryotic Genome Annotation](#)

[Prokaryotic Genome Annotation](#)

[PASC \(Pairwise Sequence Comparison\)](#)

External Resources

[GOLD - Genomes Online Database](#)

[Bacteria Genomes at Sanger](#)

[Ensembl](#)

DNA Databases and Genome Browsers



Gene

Gene integrates information from a wide range of species. A record may include nomenclature, Reference Sequences (RefSeqs), maps, pathways, variations, phenotypes, and links to genome-, phenotype-, and locus-specific resources worldwide.

Using Gene

[Gene Quick Start](#)

[FAQ](#)

[Download/FTP](#)

[RefSeq Mailing List](#)

[Gene News](#)

[Factsheet](#)

Gene Tools

[Submit GeneRIFs](#)

[Submit Correction](#)

[Statistics](#)

[BLAST](#)

[Genome Workbench](#)

[Splign](#)

Other Resources

[OMIM](#)

[RefSeq](#)

[RefSeqGene](#)

[Protein Clusters](#)

DNA Databases and Genome Browsers

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

Genomics
Institute



Genome Browser



Genomes

Genome Browser

Tools

Mirrors

Downloads

My Data

Projects

Help

About Us

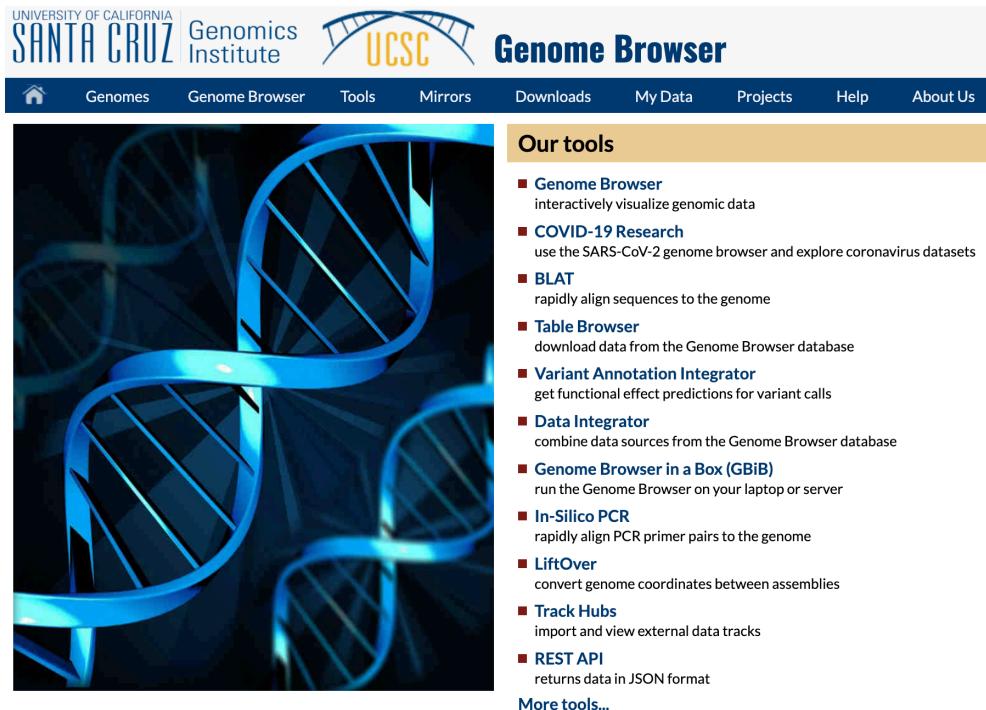


Our tools

- **Genome Browser**
interactively visualize genomic data
- **COVID-19 Research**
use the SARS-CoV-2 genome browser and explore coronavirus datasets
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBiB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **Track Hubs**
import and view external data tracks
- **REST API**
returns data in JSON format

[More tools...](#)

DNA Databases and Genome Browsers



UNIVERSITY OF CALIFORNIA | SANTA CRUZ Genomics Institute  Genome Browser

Genomes Genome Browser Tools Mirrors Downloads My Data Projects Help About Us

Our tools

- **Genome Browser**
interactively visualize genomic data
- **COVID-19 Research**
use the SARS-CoV-2 genome browser and explore coronavirus datasets
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Genome Browser in a Box (GBIB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **Track Hubs**
import and view external data tracks
- **REST API**
returns data in JSON format

[More tools...](#)

- Sequence and Annotation Downloads
- FTP
- Table Browser
- MySQL

DNA Databases and Genome Browsers

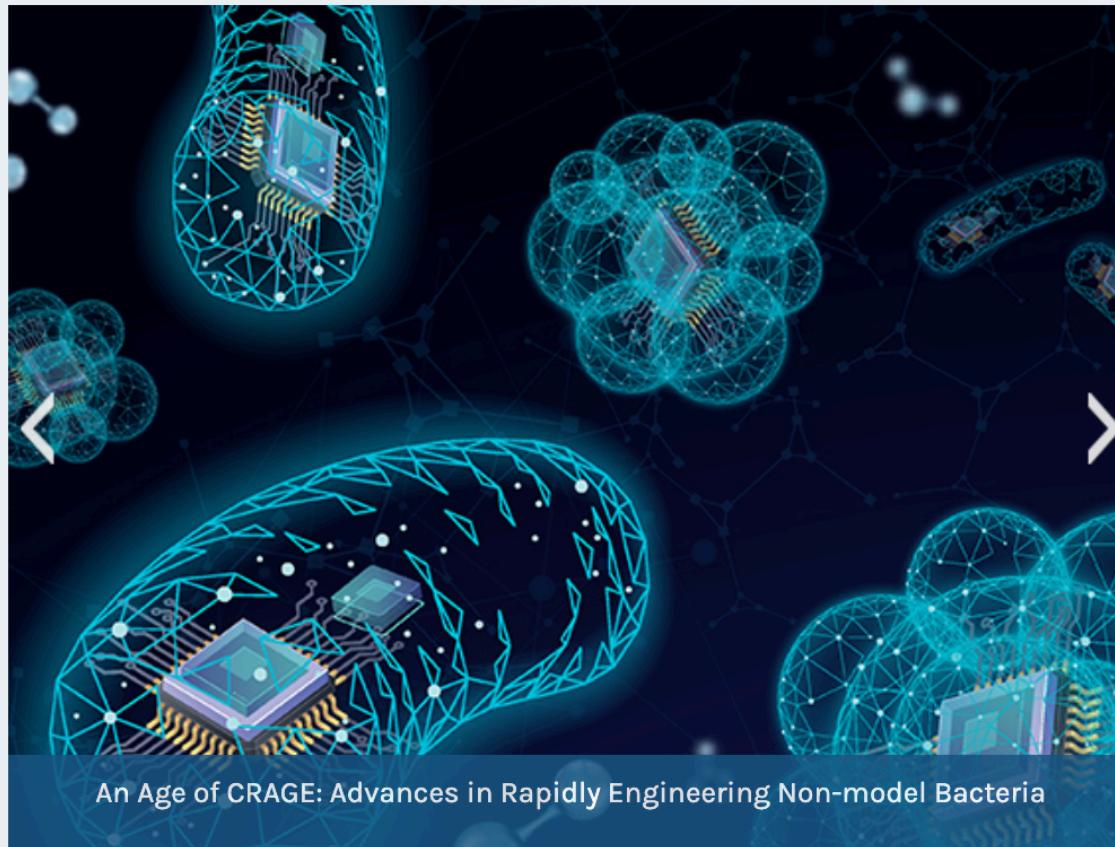
[COVID-19](#)[ABOUT US](#)[CONTACT US](#)

Search JGI websites ...

SEARCH

[Our Science](#)[Our Projects](#)[Data & Tools](#)[User Programs](#)[News & Publications](#)

A BERKELEY LAB NATIONAL USER FACILITY



Latest News

JANUARY 13, 2022

[Introducing New Members of the JGI User Executive Committee](#)

NOVEMBER 25, 2021

[Enlarging Windows into Understanding Gene Functions](#)

NOVEMBER 18, 2021

[Marine Microbe Contains Multitudes](#)

[LATEST JGI NEWS RELEASES](#)

Announcements

Letters of Intent are due March 16, 2022 to [freely access resources](#) at multiple DOE user facilities through the FICUS collaborative proposal call!

[CSP Functional Genomics proposals](#) are due January 31, 2022 for inclusion in the next biannual review.

Read our new [Data Policy](#).

DNA Databases and Genome Browsers



[COVID-19](#) [ABOUT US](#) [CONTACT US](#) Search JGI websites ... [SEARCH](#)

[Our Science](#)

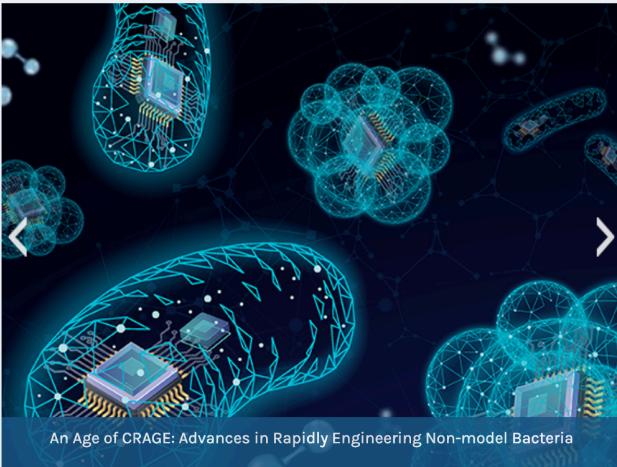
[Our Projects](#)

[Data & Tools](#)

[User Programs](#)

[News & Publications](#)

A BERKELEY LAB NATIONAL USER FACILITY



Latest News

JANUARY 13, 2022

[Introducing New Members of the JGI User Executive Committee](#)

NOVEMBER 25, 2021

[Enlarging Windows into Understanding Gene Functions](#)

NOVEMBER 18, 2021

[Marine Microbe Contains Multitudes](#)

LATEST JGI NEWS RELEASES

Announcements

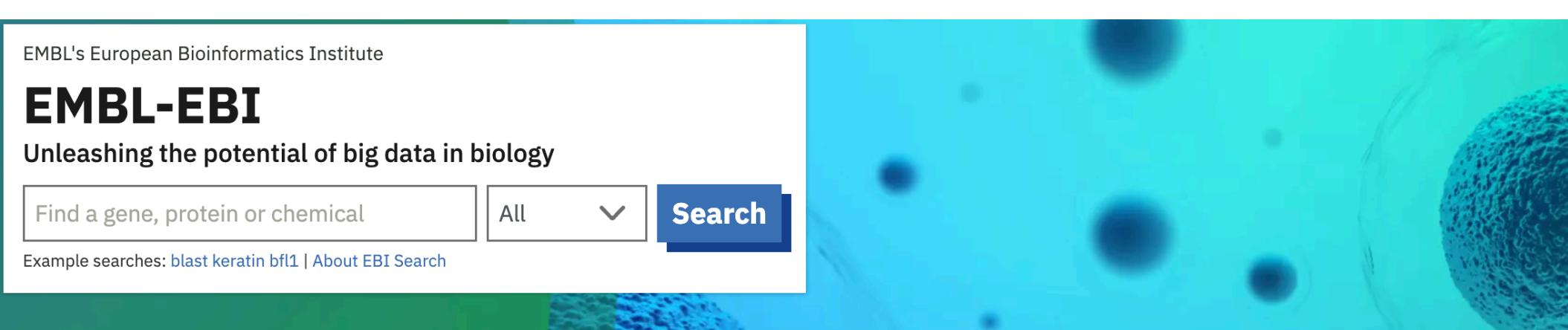
Letters of Intent are due March 16, 2022 to [freely access resources](#) at multiple DOE user facilities through the FICUS collaborative proposal call!

[CSP Functional Genomics proposals](#) are due January 31, 2022 for inclusion in the next biannual review.

Read our new [Data Policy](#)

- **Genome Portal**
- **JGI Gold**
- **Integrated Microbial Genomes and Microbiomes**
- **Integrated Microbial Genomes Content**
- **Mycocosmos**
- **Phytozome**

DNA Databases and Genome Browsers



EMBL's European Bioinformatics Institute

EMBL-EBI

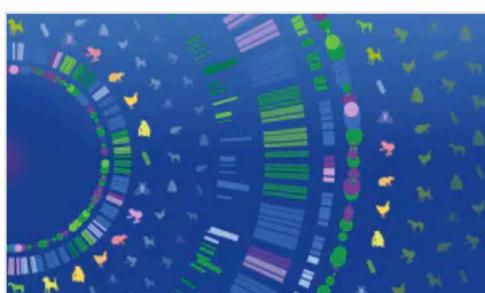
Unleashing the potential of big data in biology

Find a gene, protein or chemical All

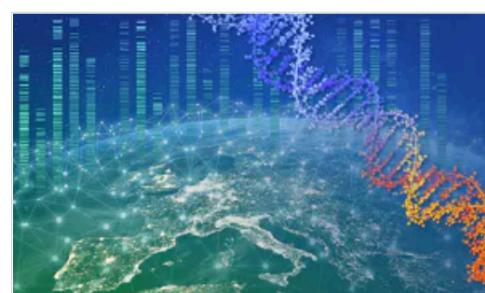
Example searches: blast keratin bfl1 | About EBI Search

[Find data resources →](#) [Submit data →](#) [Explore our research →](#) [Train with us →](#)

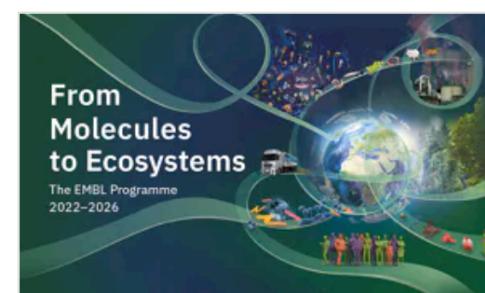
Latest news →



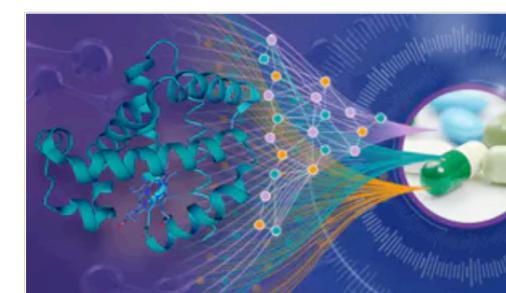
Genomic data for biodiversity – a global challenge



The Earth Biogenome Project: on track and ready to ramp up



From Molecules to Ecosystems
The EMBL Programme 2022–2026



AI for drug discovery: what's the hold up?

DNA Databases and Genome Browsers

The screenshot shows the homepage of EMBL-EBI. At the top left is the text "EMBL's European Bioinformatics Institute". Below it is the "EMBL-EBI" logo with the tagline "Unleashing the potential of big data in biology". A search bar contains the placeholder "Find a gene, protein or chemical" and a dropdown menu set to "All". To the right is a large blue button labeled "Search". Below the search bar is a link to "Example searches: blast keratin bfl1 | About EBI Search". The main navigation menu includes "Find data resources", "Submit data", "Explore our research", and "Train with us". Below the menu, there is a section titled "Latest news" with four items:

- Genomic data for biodiversity – a global challenge
- The Earth Biogenome Project: on track and ready to ramp up
- A new era for European molecular biology
- AI for drug discovery: what's the hold up?

- **EMBL-EBI-Services**
- **Uniprot**
- **Protein Data Bank**
- **Europe PMC**
- **Expression Atlas**
- **ChEMBL**

DNA Databases and Genome Browsers

Integrative Genomics Viewer



DNA Databases and Genome Browsers

Integrative Genomics Viewer

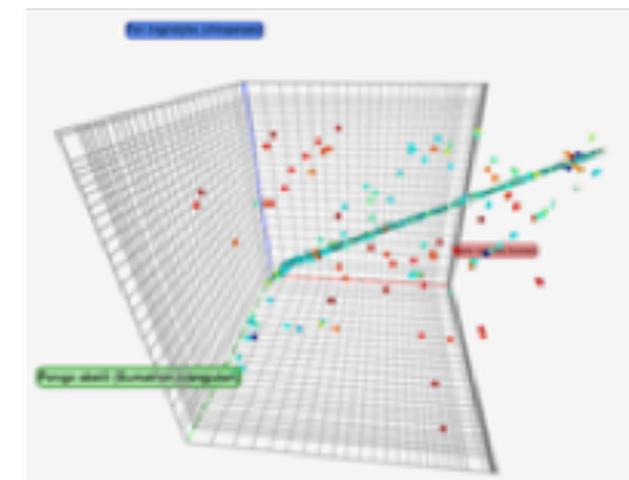
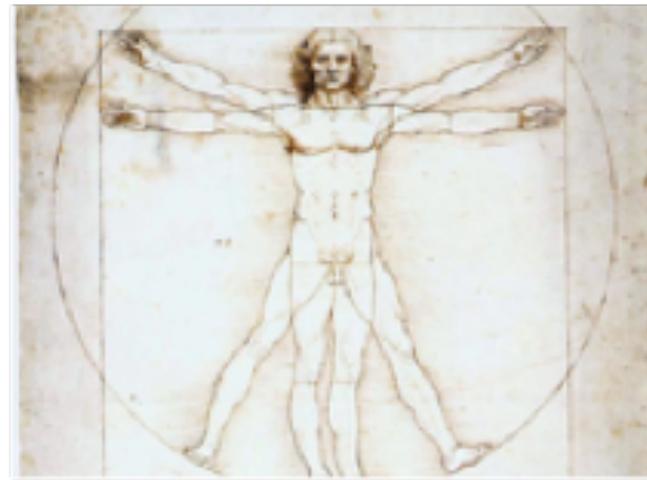
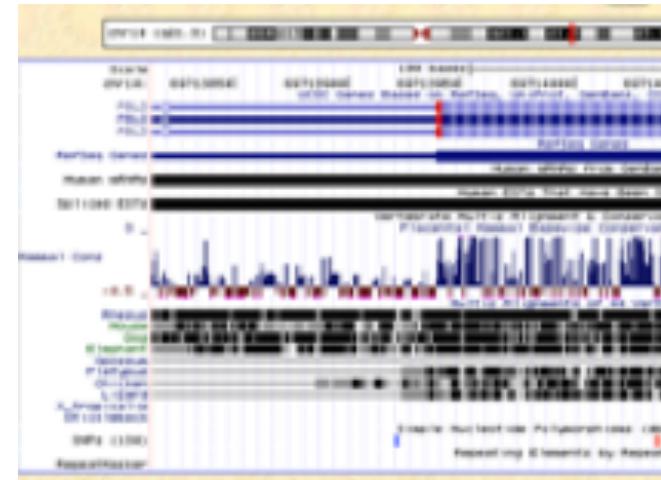


- Download IGV

Computational Genomics

Introduction To Databases

Repeats Databases



Repeats Databases

[Subscription News!](#)[Home](#)[About](#)[Register](#)[My Account](#)
[Subscribe](#)[Browse](#)[Search](#)[Repeat Masking](#)[Download](#)[Submit](#)[Repbase Reports](#)[Education](#)

GENETIC INFORMATION RESEARCH INSTITUTE

Our mission is to understand the processes altering the genetic makeup of different organisms as well as their biological and medical impact.

News

December 21, 2021

[Repbase Update Volume 26, Issue 12](#)
([Registration](#) required)

December 21, 2021

[Repbase Reports Volume 21, Issue 12](#)
(ISSN# 1534-830X; [Registration](#) required)

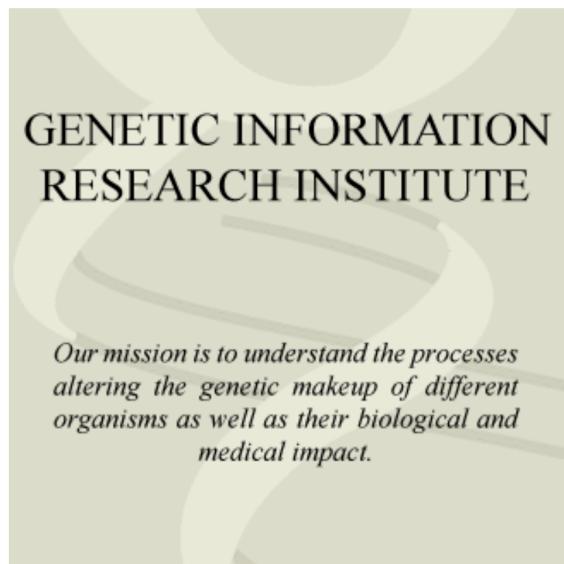
February 22, 2021

[2020 Year in Review](#). From 2017 onward, a short annual report will be presented at the beginning of each year.

[Show more news...](#)

...

DNA Databases and Genome Browsers



News

December 21, 2021

[Repbase Update Volume 26, Issue 12](#)

([Registration](#) required)

December 21, 2021

[Repbase Reports Volume 21, Issue 12](#)

(ISSN# 1534-830X; [Registration](#) required)

February 22, 2021

[2020 Year in Review](#). From 2017 onward, a short annual report will be presented at the beginning of each year.

[Show more news...](#)

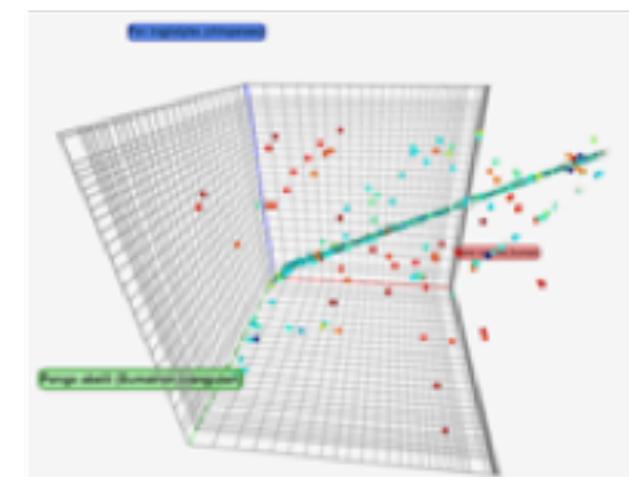
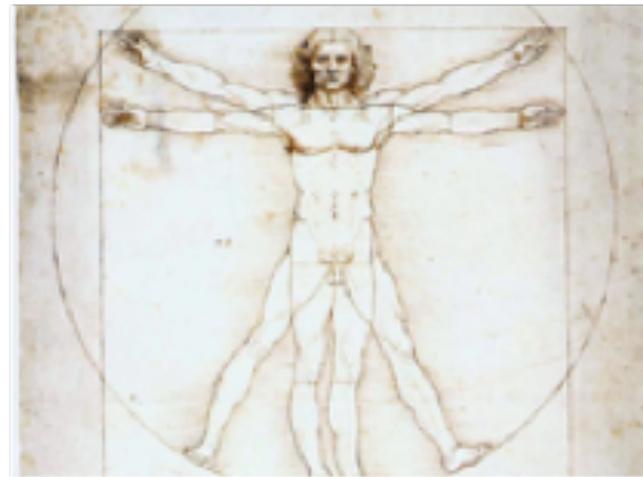
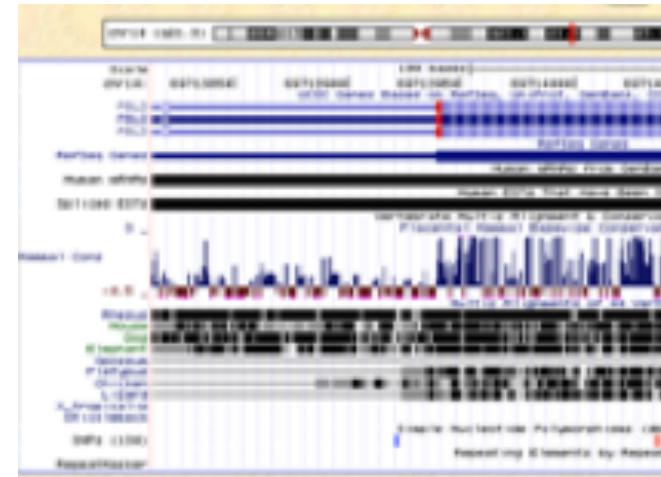
...

- **The GIRI Engine: Censor**
- **The GIRI Education Section**

Computational Genomics

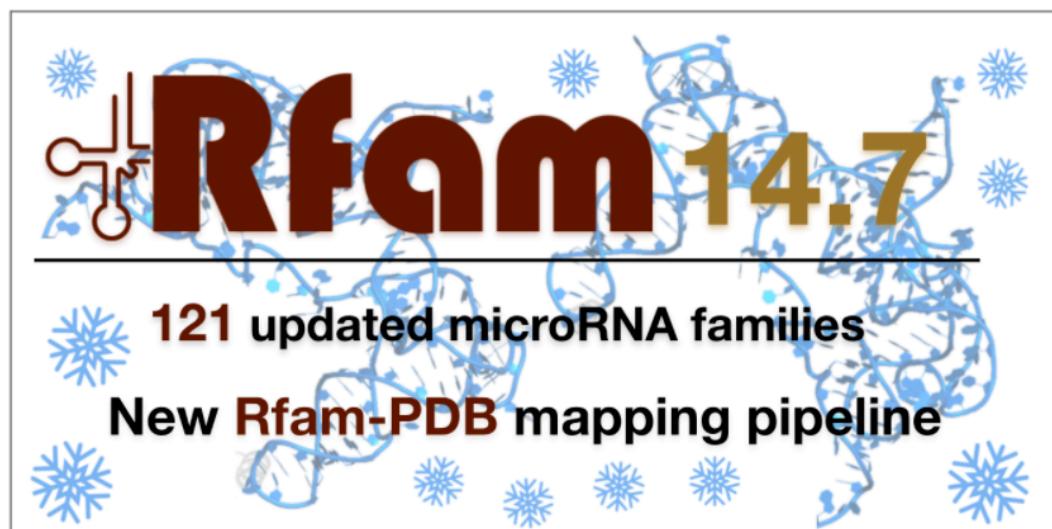
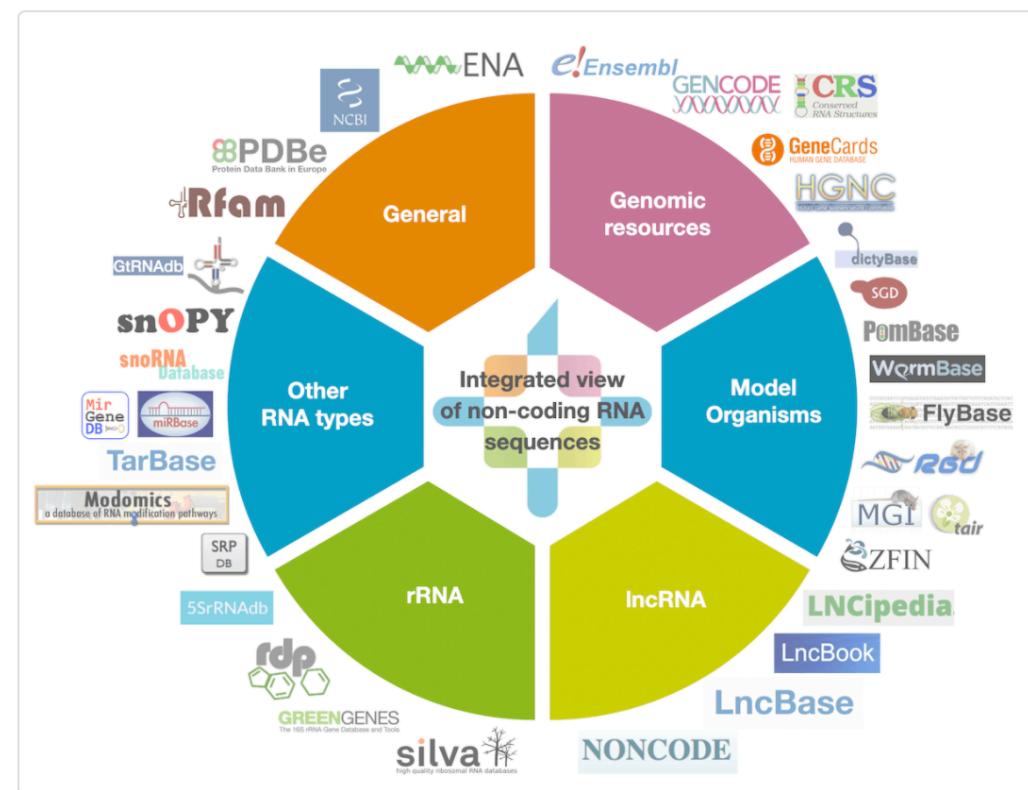
Introduction To Databases

RNA Databases



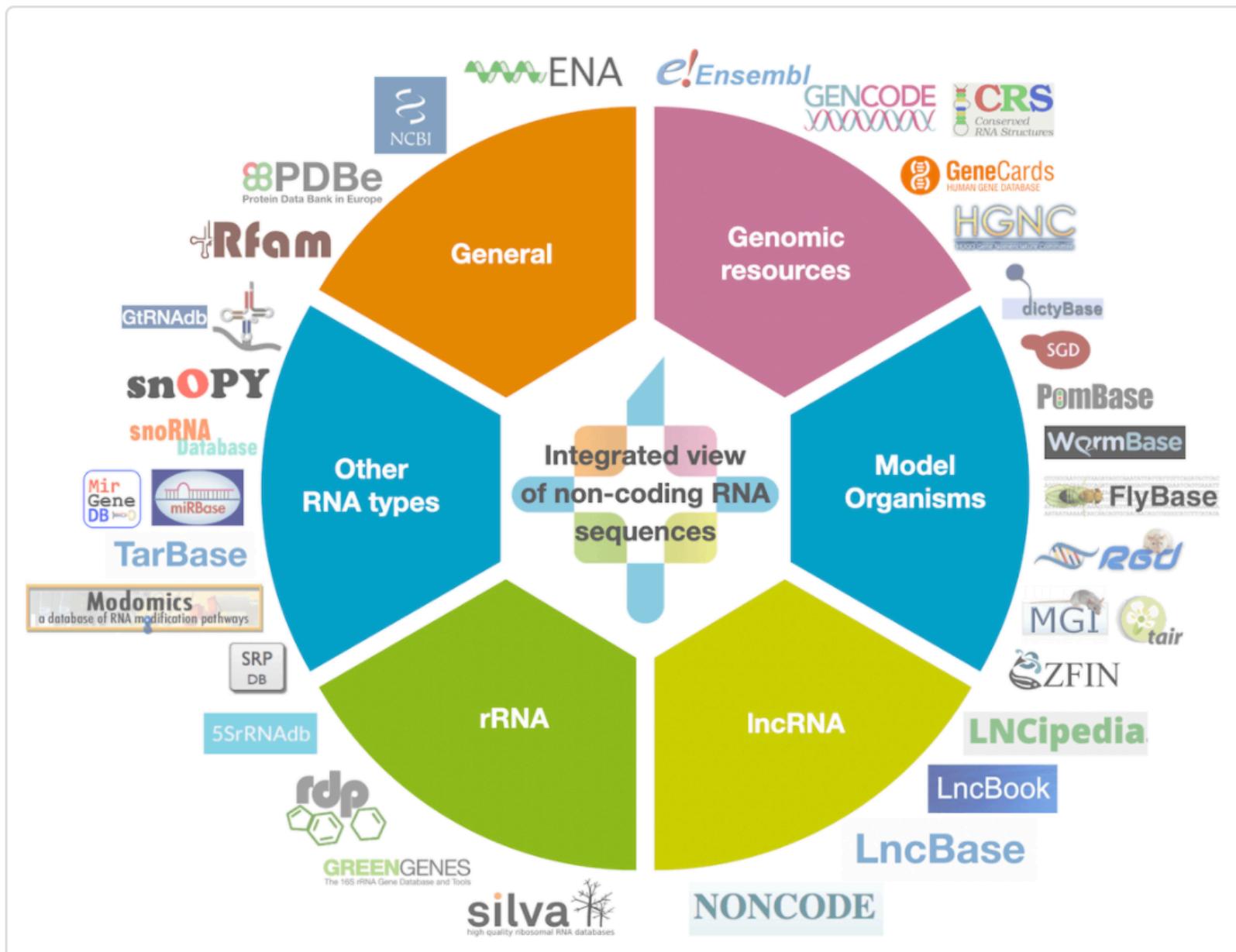
RNA Databases

RNAcentral Expert Databases



RNA Databases

RNAcentral Expert Databases



RNA Databases

RNAcentral

What is RNAcentral?

- RNAcentral is a database of ncRNA sequences of all ncRNA types from all organisms. It provides a single entry point for accessing data from numerous specialised RNA databases.
- RNAcentral provides four key functionalities:
 1. Viewing information about individual ncRNA sequences
 2. Text search that enables exploration of ncRNA sequences from different sources
 3. Sequence search for performing sequence similarity queries against a comprehensive set of ncRNA sequences
 4. FTP archive with downloadable files, including genome annotations in BED and GFF3 formats

RNA Databases

RNAcentral

Non-coding RNAs

- Similar to mRNAs, non-coding RNAs (ncRNAs) are transcribed from DNA but are not translated into proteins.
- NcRNAs are found in all organisms and have a broad range of functions. For example, protein synthesis requires transfer RNAs (tRNA) and ribosomal RNAs (rRNA), which are found in all organisms. Many ncRNAs are connected with diseases and are subject to active research.

Non-coding RNA data in RNAcentral

- RNAcentral imports most known types of non-coding RNAs, for example:
 1. tRNA
 2. rRNA
 3. microRNA
 4. lncRNA
 5. snoRNA
 6. piRNA
 7. SRP RNA
 8. vault RNA and many others

RNA Databases

RNAcentral

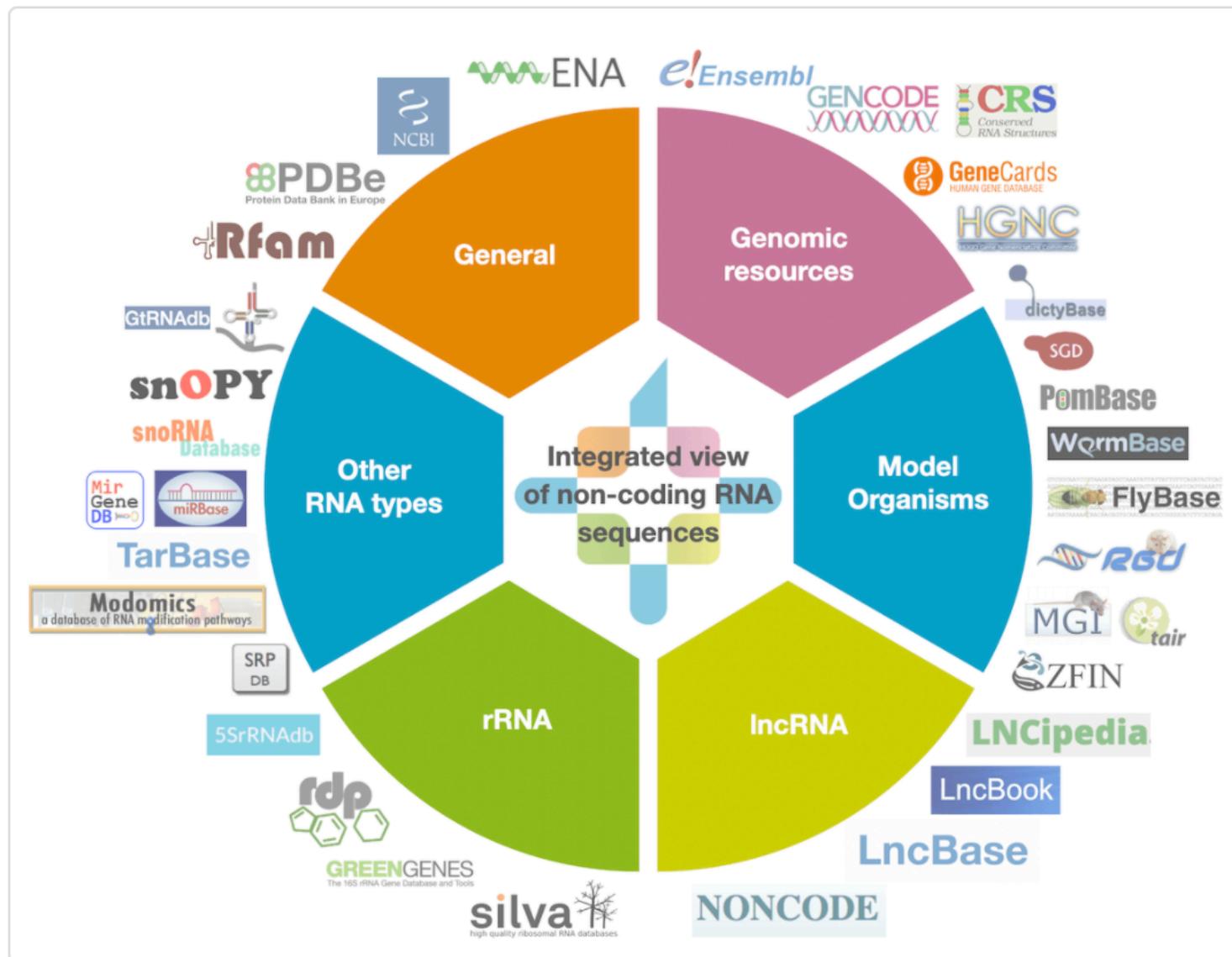
- The sequences in RNAcentral come from a wide range of species covering most of the taxonomic space.
- RNAcentral is designed as a single entry point for anyone interested in ncRNAs, where they can find a high-level overview of ncRNA content in different species, as well as functional information about individual ncRNAs. This includes genome locations, RNA secondary structure, Rfam classification, orthologs and paralogs, microRNA targets, RNA modifications, and more.
- RNAcentral not only imports the data but also generates additional annotations, such as a comprehensive genome mapping for >350 reference genomes¹ and template-based RNA secondary structure diagrams.
- RNAcentral provides four key functionalities:
 1. Viewing information about ncRNA sequences
 2. Text search that allows for exploration of ncRNAs from different member databases
 3. Sequence search for running sequence similarity queries against a comprehensive set of ncRNAs
 4. FTP archive with downloadable files in various formats, including sequences in FASTA format, genome annotations in GFF3 and BED formats, and others.
- RNAcentral contains millions of sequences and is updated every 3 months via new releases. See how the number of sequences in RNAcentral has changed over time.

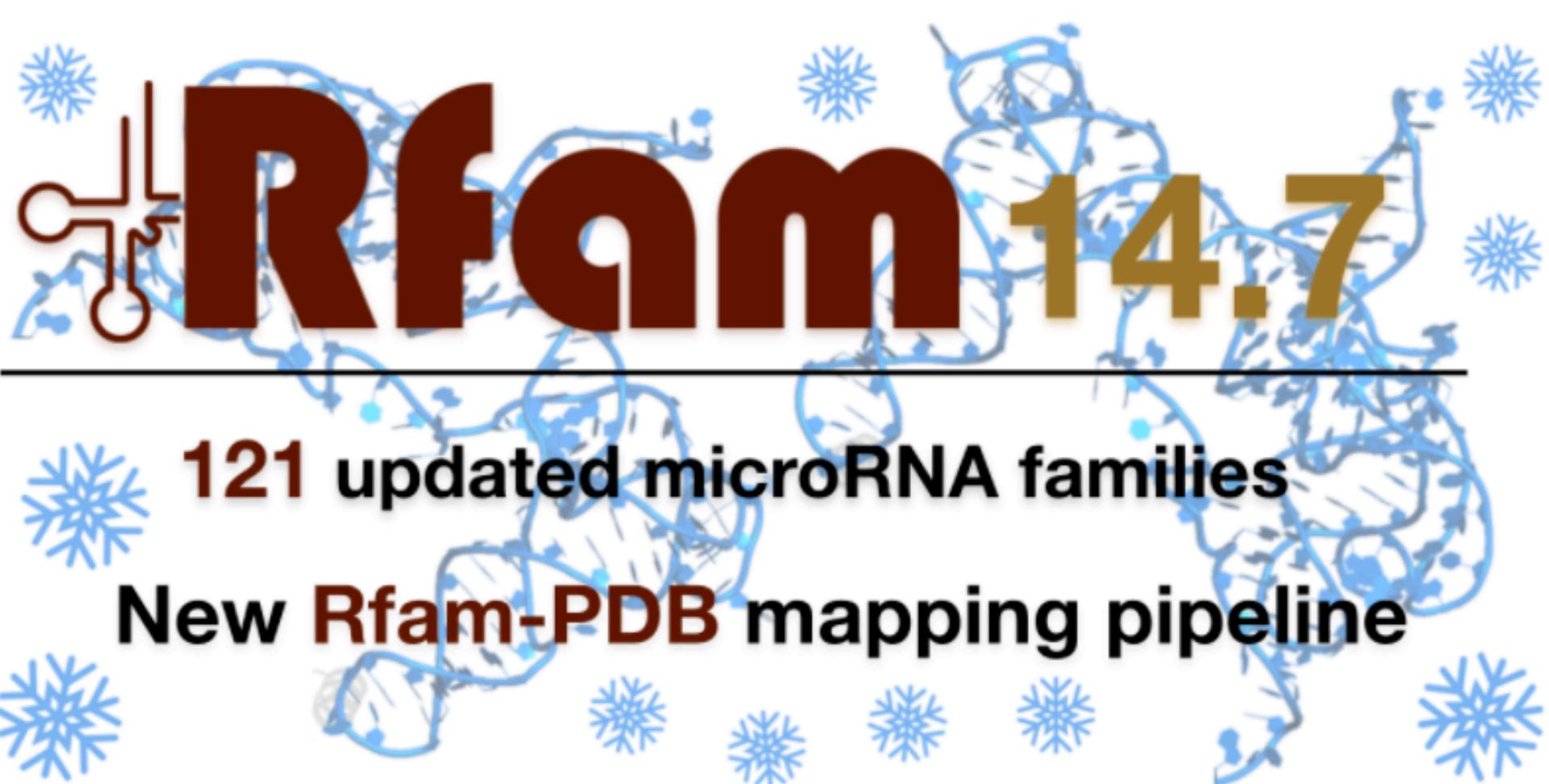
RNA Databases

RNAcentral

Where do the data come from?

RNAcentral Expert Databases





The logo features the Rfam 14.7 text in large, bold, brown and gold letters. The letter 'R' is stylized with a keyhole and a lock. The background is white with a decorative border of blue RNA helixes and snowflakes.

Rfam 14.7

121 updated microRNA families

New Rfam-PDB mapping pipeline

RNA Databases

Rfam

What is Rfam?

- Rfam is a database of non-coding RNA.
- We classify non-coding RNAs into families, using multiple sequence alignments, consensus secondary structure and covariance models. 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:
 - 1.non-coding RNA genes
 - 2.structured cis-regulatory elements
 - 3.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 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.

RNA Databases

Rfam

What is an Rfam family?

- Each family starts with a group of non-coding RNA 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 conservation. 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 (abbreviated to CM) may be thought of as more complicated hidden Markov models.

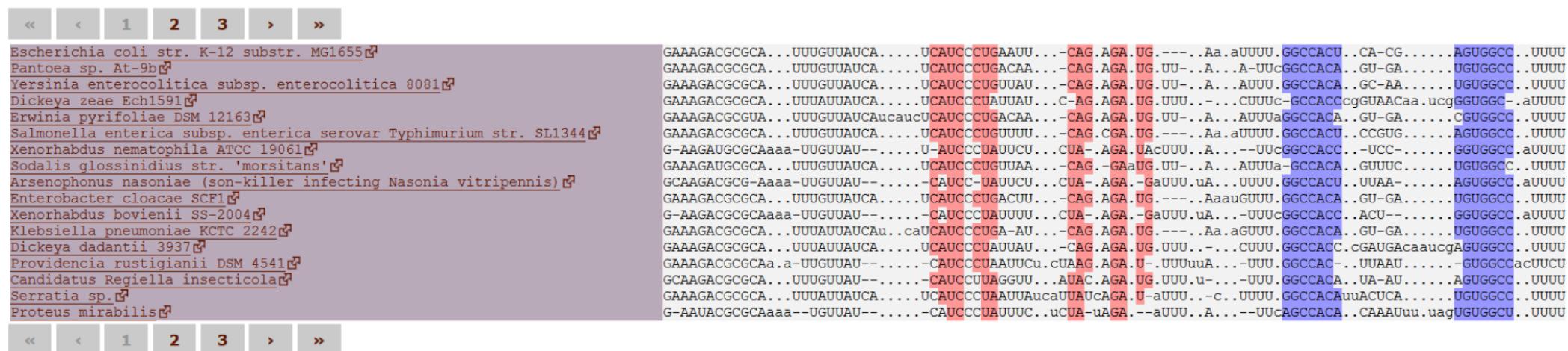
Seed sequence alignment for RF00078



RNA Databases

Rfam

Seed sequence alignment for RF00078



Example seed sequence alignment, showing the secondary structure regions (colored 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 terms for each family. Text annotation describing the family's function is provided through Wikipedia.