

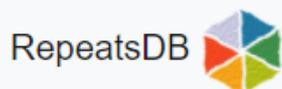


JOHANNES GUTENBERG
UNIVERSITÄT MAINZ

Examining repeats with databases

Miguel Andrade
Faculty of Biology,
Johannes Gutenberg University
Mainz, Germany
andrade@uni-mainz.de

RepeatsDB



[About](#) [Browse](#) [Search](#) [Stats](#) [Help](#)

News

RepeatsDB v3.0

Release: **2020-09-15**

[More info](#)

Start

For a fast search use the top-right search box. Alternatively, visit the browse and search pages.

Contact

For questions and/or comments, please [write to us](#) at [BioComputing UP](#).

[Follow @BiocomputingUP](#)

Citing RepeatsDB

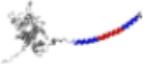
[RepeatsDB in 2021: improved data and extended](#)



RepeatsDB is a database of annotated tandem repeat protein structures. The database provides unit position, classification and reference to other databases. To start using RepeatsDB, please try the search box (top-right corner), the advanced [search](#) or the [browse](#) page.

Check our web server [RepeatsDB-lite](#) for interactive prediction of repeat units in protein structures.

RepeatsDB contains the following **repeat classes**:

Class	Description	UniProt	PDBs	PDB chains	Regions	Units	Example
1 	Crystalline aggregates	0	3	3	3	0	-
2 	Fibrous repeat	21	39	86	86	393	P04115

RepeatsDB

<https://repeatsdb.bio.unipd.it/structure/3vbpA>

PDB chain sequence

Tandem repeats shown on the PDB sequence, in alternating blue and red. Insertions are in yellow.

```
1  MGSHHHHHHENLYFQ  GHMNSFYSQEELKKI  GFLSVGKNVLIS
46  SIYNPGVISIGNNVR  IDDFCILSGKVTIGS  YSHIAAYTALYC
91  GIEMYDFANISSRTI  VYAAINDFSGNALMG  PTIPNQYKNVKT
136 ILKKHVIIGAHSIIF  PNVVIGEGVAVGAMS  MVKESLDDWYIV
181 PVRKIKARKRKIVEL  ENEFLKSMNS
```

PDB Structure

Tandem repeats shown on the PDB structure, in alternating blue and red. Insertions are in yellow.

Multiple structural alignment

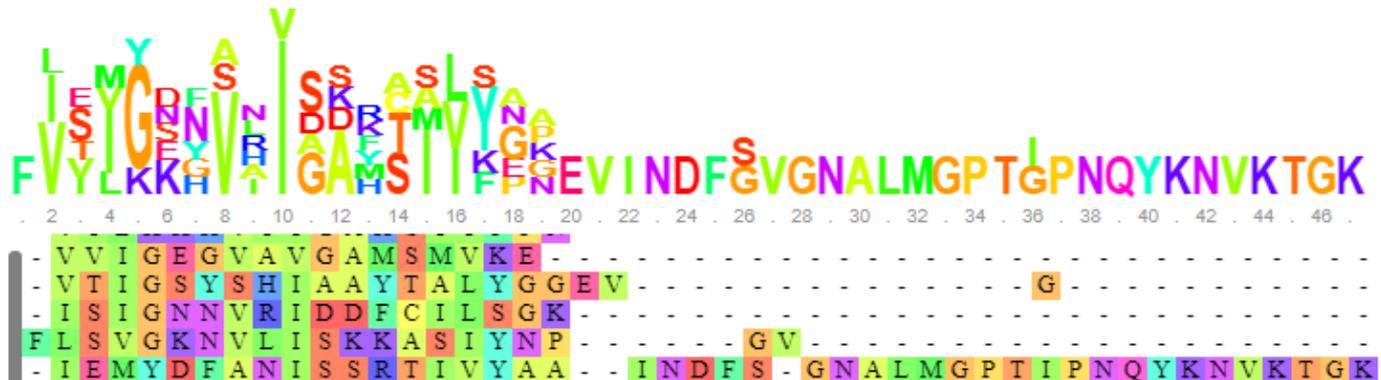
[Download PDB file](#)

The Viewer shows the structural overlapping of units.

Multiple sequence alignment

[Download FASTA file](#)

The Viewer shows the sequence overlapping of units.



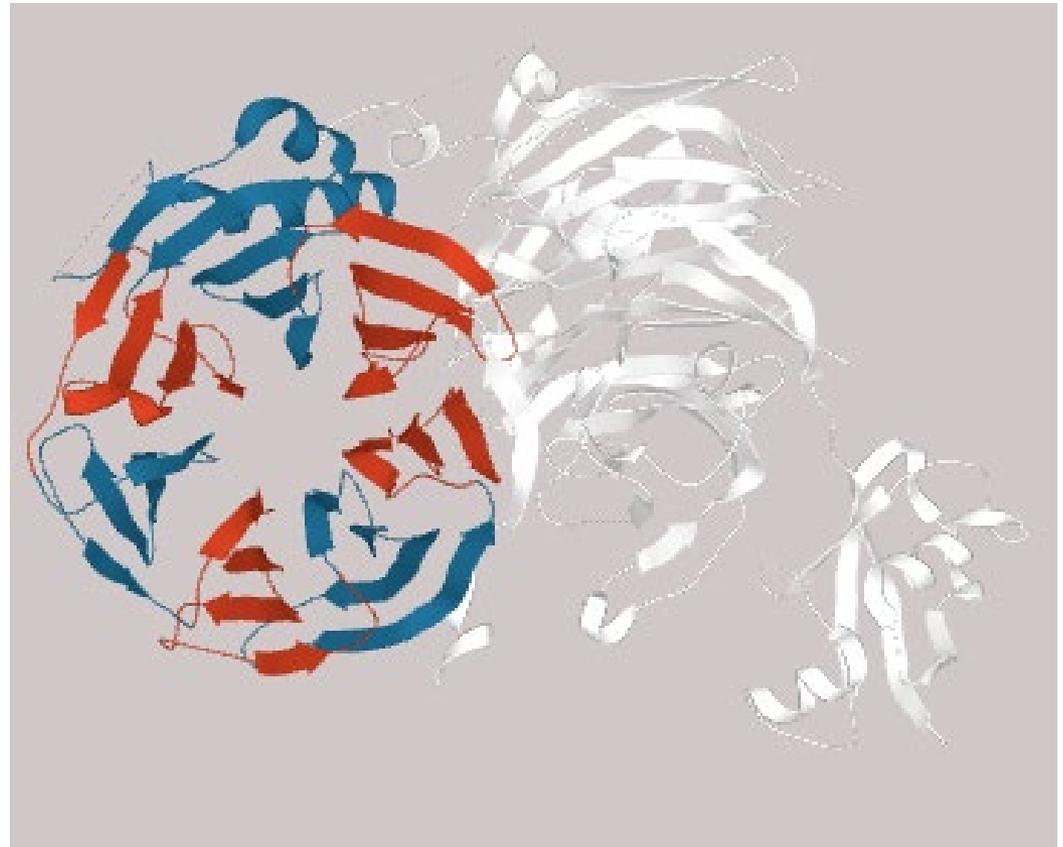
Exercise 1. Examining repeat structures in repeatsDB

- Go to RepeatsDB: <http://repeatsdb.bio.unipd.it/>
- Click the "Search" tab, then search "Repeat unit number"
- Choose one example (you might look for an unreviewed one – white star) and **write down the PDB name (e.g. 6yc3B)**.
- Check the assignment of the units. Is it correct? What about insertions? Are there mistakes? Write an evaluation: Looks good / one repeat wrong / many mistakes / total mess**

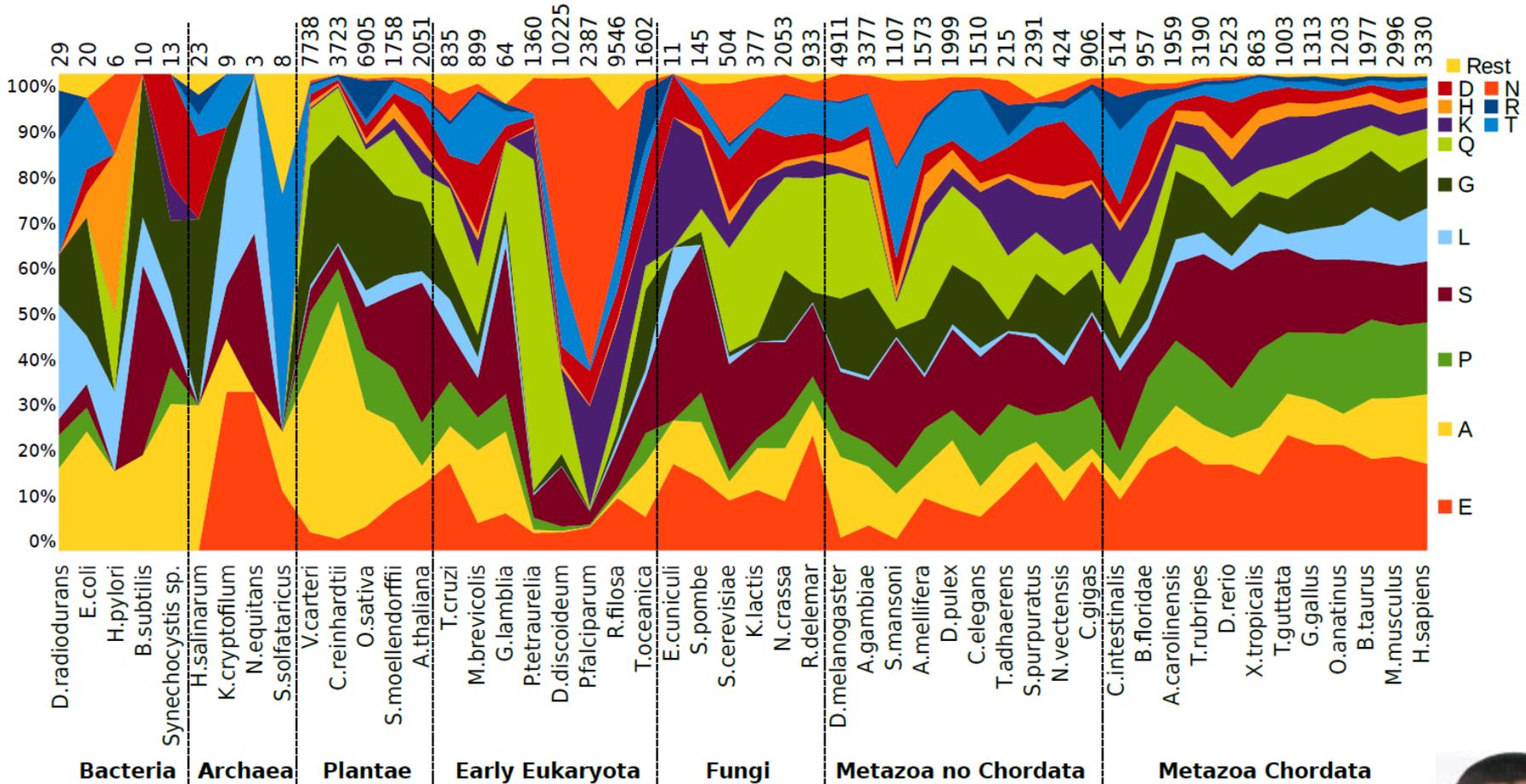
Example

<https://repeatsdb.bio.unipd.it/structure/5cxbB>

1. looks good
2. one repeat wrong
- 3. many mistakes**
4. total mess



Evolution of homorepeats in 50 species



Pablo Mier



Context and evolution of homorepeats

Input & output overview

Query: ENSP00000346389 Time elapsed: 1 seconds Clustered results: [Download](#) Ortholog's multifasta file: [Download](#) Ortholog's alignment (fasta format): [Download](#)

Cutoff6: - Cutoff8: - Cutoff10: - Cutoff12: - Cutoff14: - Cutoff16: - Cutoff18: - Cutoff20: - Cutoff22: -

EnsemblProtein ID: ENSP00000346389
AC: Q02078
UniProtID: MEF2A_HUMAN
Description: Myocyte-specific enhancer factor 2A (Serum response factor-like protein 1)
Gene Name: MEF2A MEF2



dAtabase of PolyX Evolution

Mier *et al.* (2016) Bioinformatics

Exercise 2. Viewing homorepeats in an alignment with dAPE

- Find a random human protein in **UniProt** using this URL:
<https://www.uniprot.org/uniprot/?query=organism%3A%22Homo+sapiens+%28Human%29+%5B9606%5D%22+reviewed%3Ayes&random=yes>. **Write the Entry Name.**
- Copy the Entry ID (e.g. P10275).
- Go to the dAPE web page:
<http://cbdm-01.zdv.uni-mainz.de/~munoz/polyx/>
- Use the Entry ID in option A
- Hit the "Report the evolution of its polyX" button
- Write down: (1) Which type of polyX has the human protein?
(2) Which **additional** polyX not in the human protein were found in the orthologs?

Exercise 3. Find structures of short repeats

- Go to the PDB web page: <https://www.rcsb.org/>
- Go to Search > Advanced Search
- Type a short repeat in the Sequence Motif: as many repeats as possible; Use Mode PROSITE and RegExp (example: QQQQ / another example: [SP]SA[SP]SA[SP]SA)
- Click “count” to see if there are any cases
- Check the structure (3D View) to see if the repeat is in the structure (it could be disordered = absent)
- Write down the PDB name, RegExp and structure
Example 3NB9 QQQQ alpha-helix
Example 3FRC NNNN absent