BGSU-NDB Collaboration: RNA Databases Update

Neocles Leontis
Benasque
20 July 2018

RNA.bgsu.edu
ndbserver.rutgers.edu/

Acknowledgements

Craig Zirbel

Sri Appasamy

Jamie Cannone

Dr. Anton Petrov

Dr. Blake Sweeney

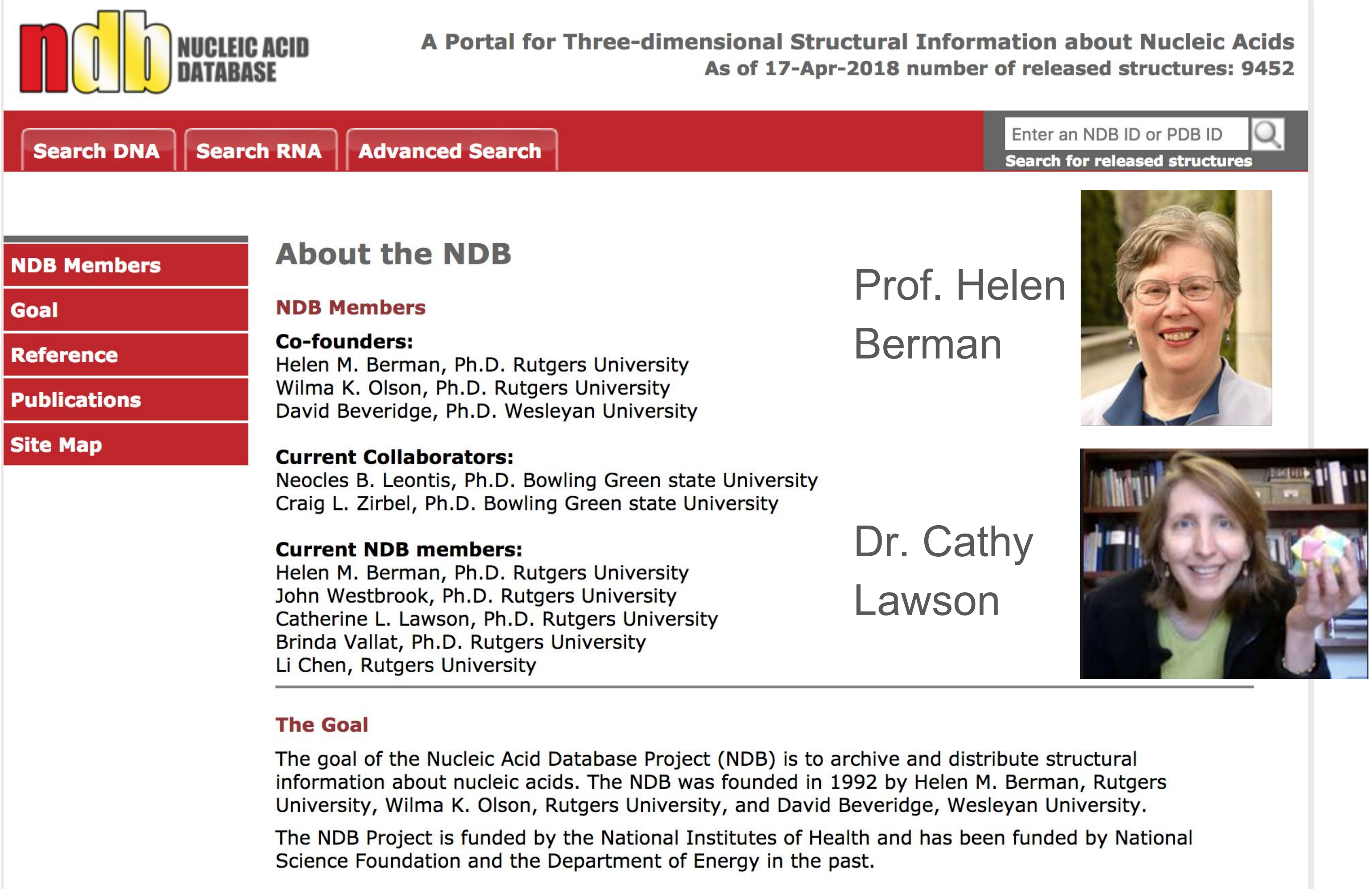
BGSU®

ndb NUCLEIC ACID DATABASE

Funding:

NIH

Collaboration with Nucleic Acid Database



A Portal for Three-dimensional Structural Information about Nucleic Acids
As of 17-Apr-2018 number of released structures: 9452

About the NDB

NDB Members

Co-founders:
Helen M. Berman, Ph.D. Rutgers University
Wilma K. Olson, Ph.D. Rutgers University
David Beveridge, Ph.D. Wesleyan University

Current Collaborators:
Neocles B. Leontis, Ph.D. Bowling Green state University
Craig L. Zirbel, Ph.D. Bowling Green state University

Current NDB members:
Helen M. Berman, Ph.D. Rutgers University
John Westbrook, Ph.D. Rutgers University
Catherine L. Lawson, Ph.D. Rutgers University
Brinda Vallat, Ph.D. Rutgers University
Li Chen, Rutgers University

Prof. Helen Berman

Dr. Cathy Lawson

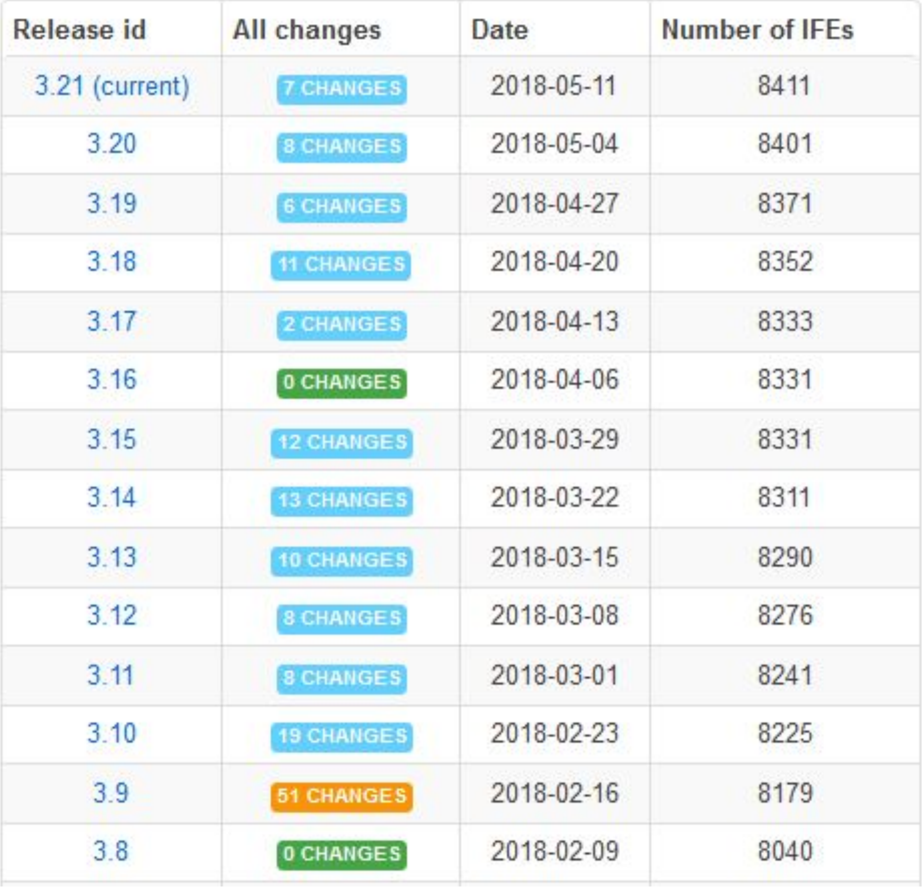
The Goal

The goal of the Nucleic Acid Database Project (NDB) is to archive and distribute structural information about nucleic acids. The NDB was founded in 1992 by Helen M. Berman, Rutgers University, Wilma K. Olson, Rutgers University, and David Beveridge, Wesleyan University.

The NDB Project is funded by the National Institutes of Health and has been funded by National Science Foundation and the Department of Energy in the past.

Weekly annotation pipeline run on all RNA-containing 3D structures in PDB:

1. Grouping structures based on Seq and 3D similarity
2. Annotation of RNA Interactions
3. Identification of 3D Motifs



Release id	All changes	Date	Number of IFEs
3.21 (current)	7 CHANGES	2018-05-11	8411
3.20	8 CHANGES	2018-05-04	8401
3.19	6 CHANGES	2018-04-27	8371
3.18	11 CHANGES	2018-04-20	8352
3.17	2 CHANGES	2018-04-13	8333
3.16	0 CHANGES	2018-04-06	8331
3.15	12 CHANGES	2018-03-29	8331
3.14	13 CHANGES	2018-03-22	8311
3.13	10 CHANGES	2018-03-15	8290
3.12	8 CHANGES	2018-03-08	8276
3.11	8 CHANGES	2018-03-01	8241
3.10	19 CHANGES	2018-02-23	8225
3.9	51 CHANGES	2018-02-16	8179
3.8	0 CHANGES	2018-02-09	8040

rna.bgsu.edu Representative sets

… running since 2011

Now based on mmCIF files:

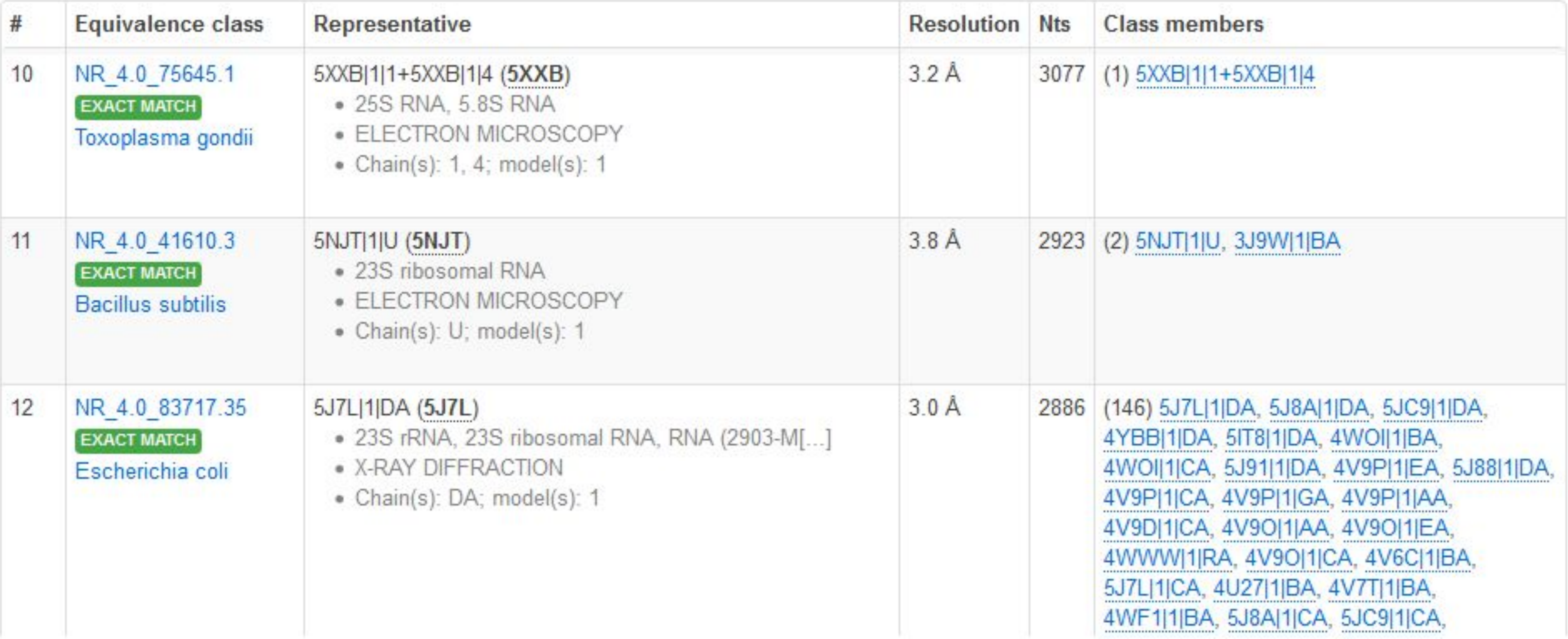
IFE = Integrated Functional Element: All RNA chains that form 2D structure together

Ex: Eukaryal LSU rRNA: 28S+5.8S

Release id	All changes	Date	Number of IFEs
0.15		2011-04-30	1699
0.14	4 CHANGES	2011-04-23	1699
0.13		2011-04-16	1696
0.12	3 CHANGES	2011-04-11	1696
0.11	12 CHANGES	2011-04-09	1696
0.10		2011-04-02	1678
0.9	3 CHANGES	2011-03-26	1677
0.8		2011-03-19	1670
0.7	3 CHANGES	2011-03-12	1670
0.6		2011-03-05	1652
0.5	2 CHANGES	2011-02-26	1650
0.4		2011-02-19	1642
0.3		2011-02-16	1642
0.2	3 CHANGES	2011-02-12	1639
0.1		2011-02-05	1630

rna.bgsu.edu Representative sets

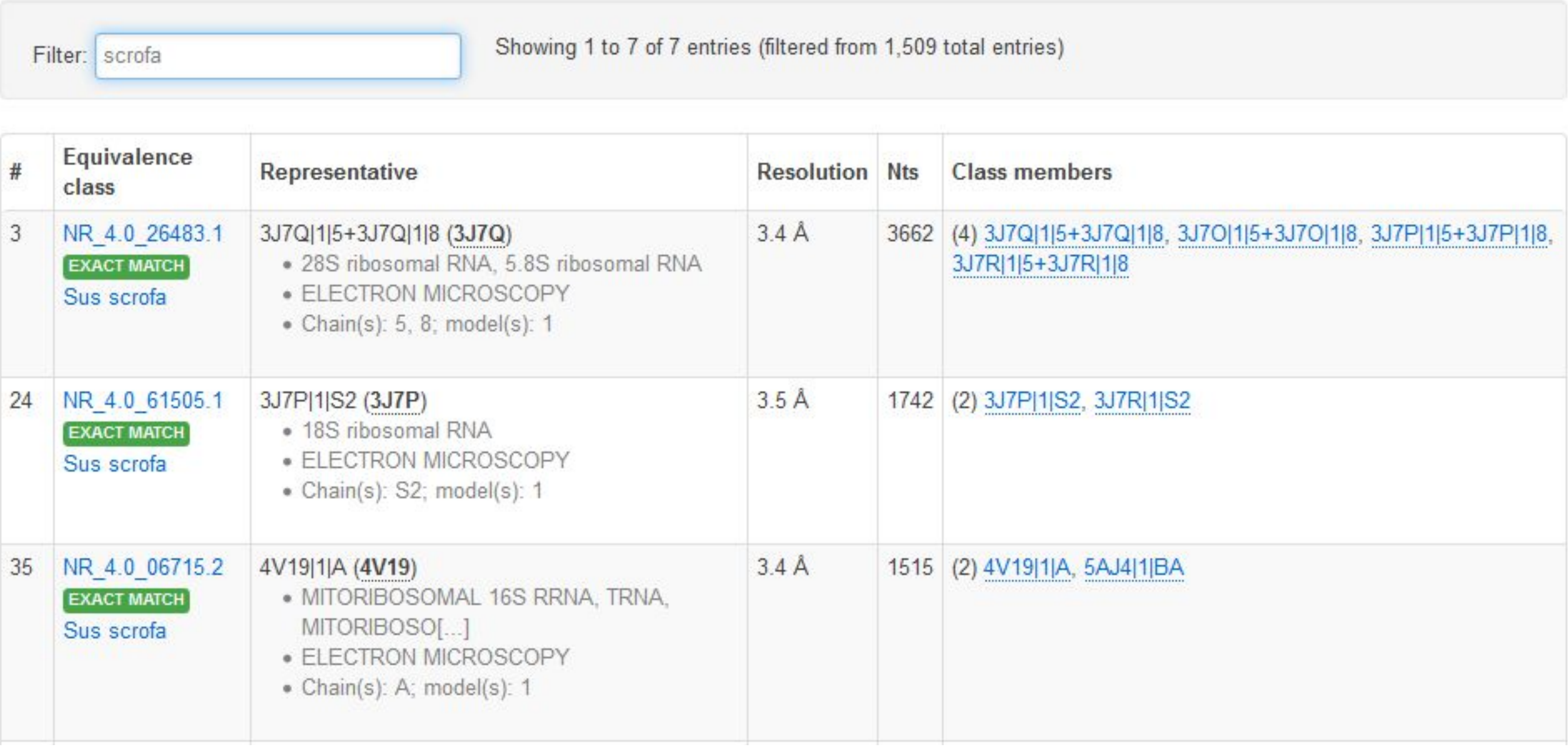
Equivalence classes - same molecule, same organism



#	Equivalence class	Representative	Resolution	Nts	Class members
10	NR_4.0_75645.1 EXACT MATCH Toxoplasma gondii	5XXB|1|1+5XXB|1|4 (**5XXB**) • 25S RNA, 5.8S RNA • ELECTRON MICROSCOPY • Chain(s): 1, 4; model(s): 1	3.2 Å	3077	(1) 5XXB|1|1+5XXB|1|4
11	NR_4.0_41610.3 EXACT MATCH Bacillus subtilis	5NJT|1|U (**5NJT**) • 23S ribosomal RNA • ELECTRON MICROSCOPY • Chain(s): U; model(s): 1	3.8 Å	2923	(2) 5NJT|1|U, 3J9W|1|BA
12	NR_4.0_83717.35 EXACT MATCH Escherichia coli	5J7L|1|DA (**5J7L**) • 23S rRNA, 23S ribosomal RNA, RNA (2903-M[...] • X-RAY DIFFRACTION • Chain(s): DA; model(s): 1	3.0 Å	2886	(146) 5J7L|1|DA, 5J8A|1|DA, 5JC9|1|DA, 4YBB|1|DA, 5IT8|1|DA, 4WOI|1|BA, 4WOI|1|CA, 5J91|1|DA, 4V9P|1|EA, 5J88|1|DA, 4V9P|1|CA, 4V9P|1|GA, 4V9P|1|AA, 4V9D|1|CA, 4V9O|1|AA, 4V9O|1|EA, 4WWW|1|RA, 4V9O|1|CA, 4V6C|1|BA, 5J7L|1|CA, 4U27|1|BA, 4V7T|1|BA, 4WF1|1|BA, 5J8A|1|CA, 5JC9|1|CA,

Where? rna.bgsu.edu/rna3dhub/nrlist/release/3.21

Filter by organism; SSU's separate from LSU's



Filter: scrofa

Showing 1 to 7 of 7 entries (filtered from 1,509 total entries)

#	Equivalence class	Representative	Resolution	Nts	Class members
3	NR_4.0_26483.1 EXACT MATCH Sus scrofa	3J7Q|1|5+3J7Q|1|8 (3J7Q) • 28S ribosomal RNA, 5.8S ribosomal RNA • ELECTRON MICROSCOPY • Chain(s): 5, 8; model(s): 1	3.4 Å	3662	(4) 3J7Q|1|5+3J7Q|1|8, 3J7O|1|5+3J7O|1|8, 3J7P|1|5+3J7P|1|8, 3J7R|1|5+3J7R|1|8
24	NR_4.0_61505.1 EXACT MATCH Sus scrofa	3J7P|1|S2 (3J7P) • 18S ribosomal RNA • ELECTRON MICROSCOPY • Chain(s): S2; model(s): 1	3.5 Å	1742	(2) 3J7P|1|S2, 3J7R|1|S2
35	NR_4.0_06715.2 EXACT MATCH Sus scrofa	4V19|1|A (4V19) • MITORIBOSOMAL 16S RRNA, TRNA, MITORIBOSO[...] • ELECTRON MICROSCOPY • Chain(s): A; model(s): 1	3.4 Å	1515	(2) 4V19|1|A, 5AJ4|1|BA

Where? rna.bgsu.edu/rna3dhub/nrlist/release/3.21

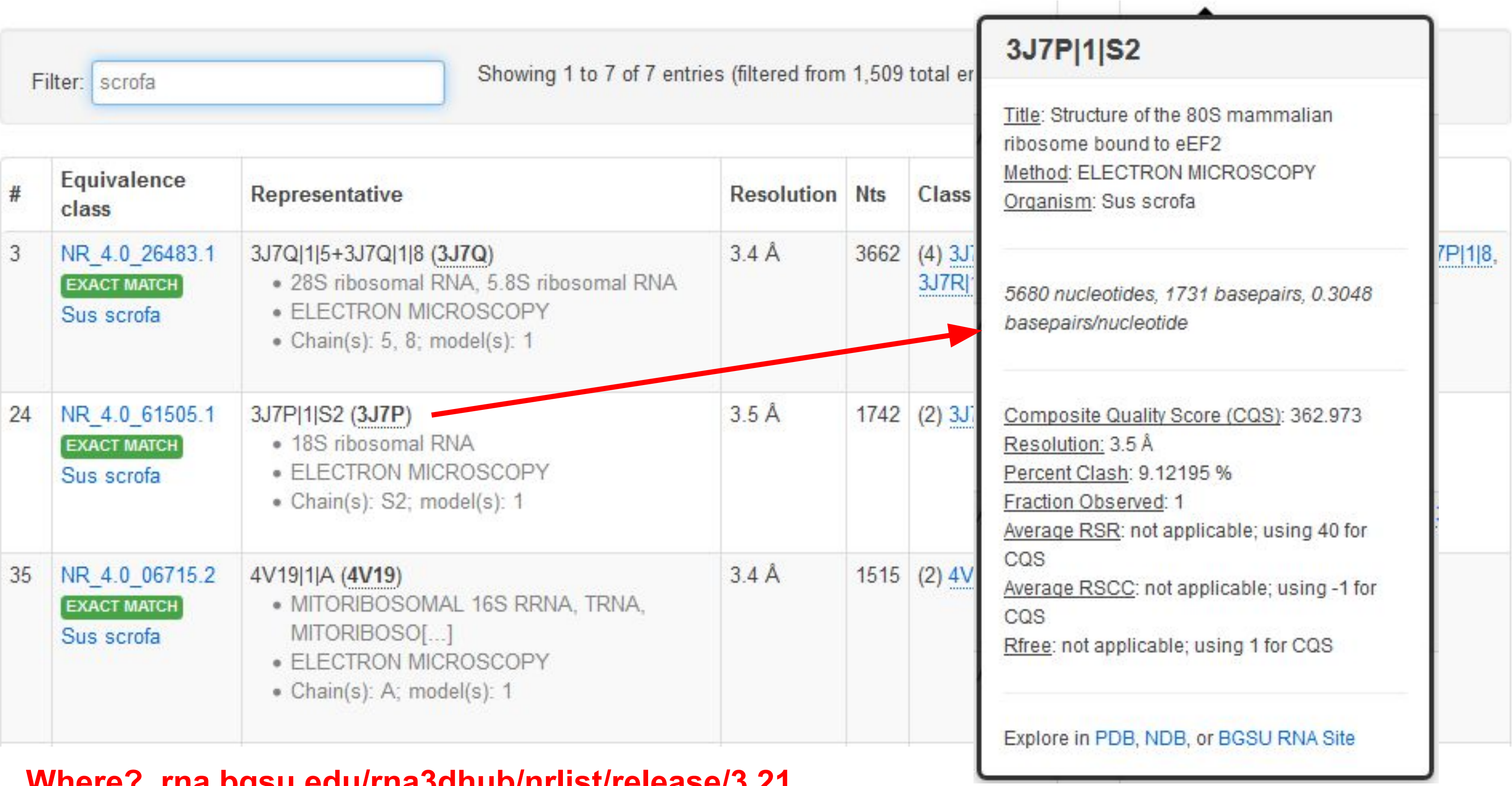
Composite Quality Score (CQS)

The representative for each equivalence class (EC) is chosen, starting with Release 3.0, as the Integrated Functional Element (IFE) with the optimal composite quality score (CQS). The CQS is the weighted sum of six components; five are defined at the IFE level, and one (Fraction Unobserved) is a property of the EC. The weights were developed to select the structure with the best experimental evidence for its reported coordinates as the EC representative.

CQS = (Resolution * 1.0) + (Percent Clash * 0.6) + (Average RSR * 8.0) + ((1 - Average RSCC) * 8.0) + (R_{free} * 18.0) + (Fraction Unobserved * 4.0)

[Percent Clash is the percentage of nucleotides having steric clashes; Fraction Unobserved is the fraction of nucleotides in the molecule which are not observed in the structure, compared to the longest experimental sequence in the EC.]

Popover shows structure quality details



Filter: scrofa

Showing 1 to 7 of 7 entries (filtered from 1,509 total er

#	Equivalence class	Representative	Resolution	Nts	Class	
3	NR_4.0_26483.1 EXACT MATCH Sus scrofa	3J7Q|1|5+3J7Q|1|8 (3J7Q) • 28S ribosomal RNA, 5.8S ribosomal RNA • ELECTRON MICROSCOPY • Chain(s): 5, 8; model(s): 1	3.4 Å	3662	(4) 3J 3J7R|	7P|1|8,
24	NR_4.0_61505.1 EXACT MATCH Sus scrofa	3J7P|1|S2 (3J7P) • 18S ribosomal RNA • ELECTRON MICROSCOPY • Chain(s): S2; model(s): 1	3.5 Å	1742	(2) 3J	
35	NR_4.0_06715.2 EXACT MATCH Sus scrofa	4V19|1|A (4V19) • MITORIBOSOMAL 16S RRNA, TRNA, MITORIBOSO[...] • ELECTRON MICROSCOPY • Chain(s): A; model(s): 1	3.4 Å	1515	(2) 4V	

3J7P|1|S2

Title: Structure of the 80S mammalian ribosome bound to eEF2
Method: ELECTRON MICROSCOPY
Organism: Sus scrofa

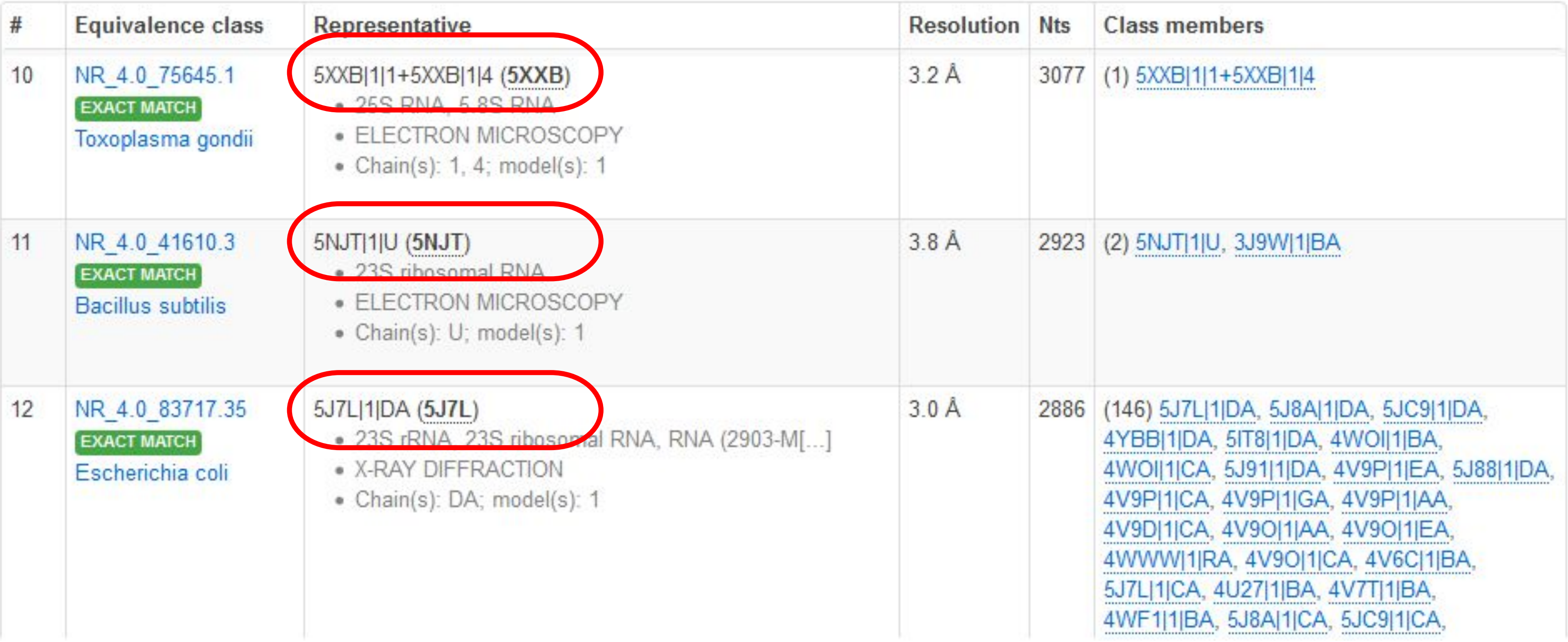
5680 nucleotides, 1731 basepairs, 0.3048 basepairs/nucleotide

Composite Quality Score (CQS): 362.973
Resolution: 3.5 Å
Percent Clash: 9.12195 %
Fraction Observed: 1
Average RSR: not applicable; using 40 for CQS
Average RSCC: not applicable; using -1 for CQS
Rfree: not applicable; using 1 for CQS

Explore in PDB, NDB, or BGSU RNA Site

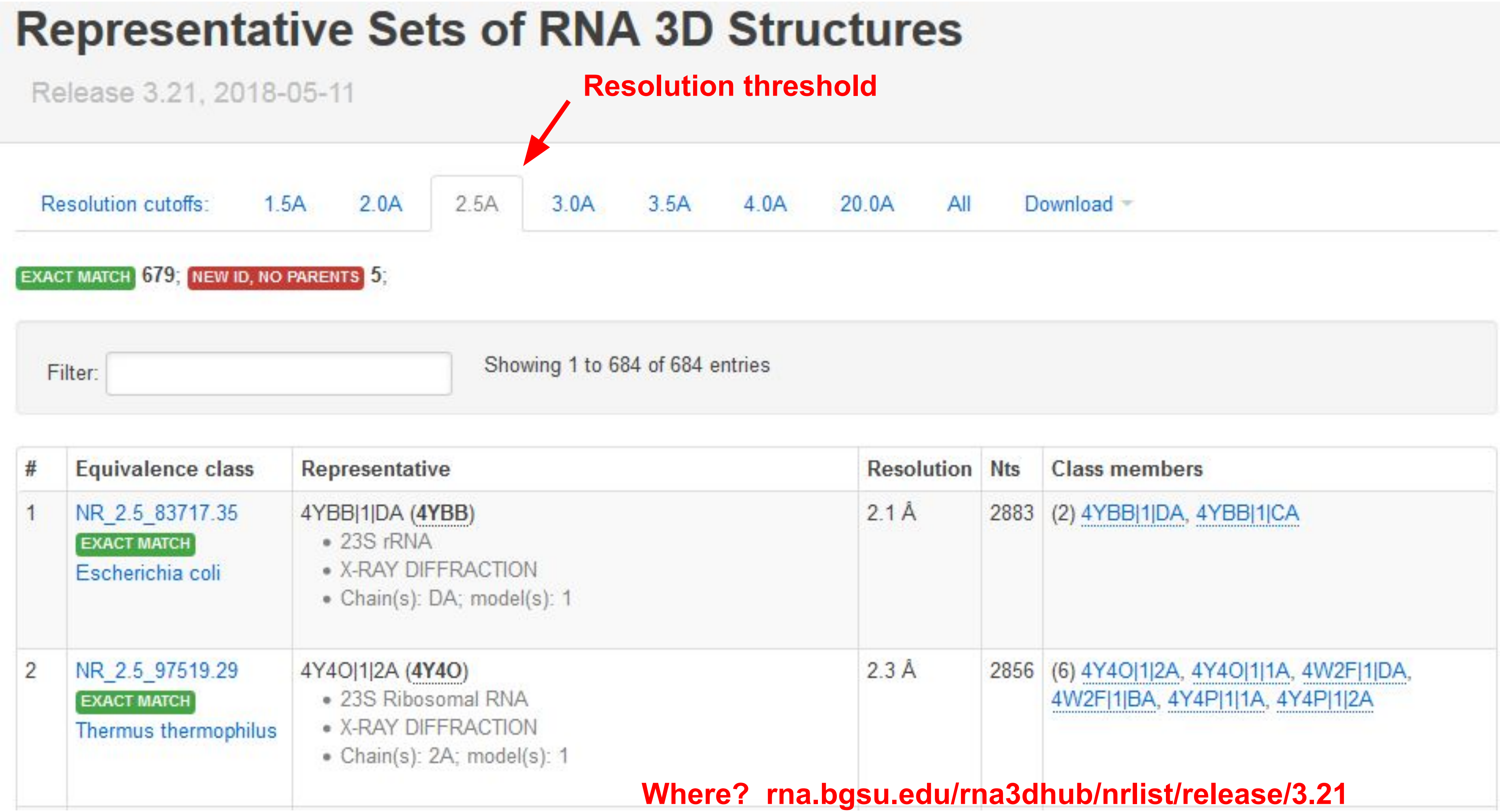
Where? rna.bgsu.edu/rna3dhub/nrlist/release/3.21

Representative set - take one IFE from each class



#	Equivalence class	Representative	Resolution	Nts	Class members
10	NR_4.0_75645.1 EXACT MATCH Toxoplasma gondii	5XXB|1|1+5XXB|1|4 (**5XXB**) • 25S RNA, 5.8S RNA • ELECTRON MICROSCOPY • Chain(s): 1, 4; model(s): 1	3.2 Å	3077	(1) 5XXB|1|1+5XXB|1|4
11	NR_4.0_41610.3 EXACT MATCH Bacillus subtilis	5NJT|1|U (**5NJT**) • 23S ribosomal RNA • ELECTRON MICROSCOPY • Chain(s): U; model(s): 1	3.8 Å	2923	(2) 5NJT|1|U, 3J9W|1|BA
12	NR_4.0_83717.35 EXACT MATCH Escherichia coli	5J7L|1|DA (**5J7L**) • 23S rRNA, 23S ribosomal RNA, RNA (2903-M[...] • X-RAY DIFFRACTION • Chain(s): DA; model(s): 1	3.0 Å	2886	(146) 5J7L|1|DA, 5J8A|1|DA, 5JC9|1|DA, 4YBB|1|DA, 5IT8|1|DA, 4WOI|1|BA, 4WOI|1|CA, 5J91|1|DA, 4V9P|1|EA, 5J88|1|DA, 4V9P|1|CA, 4V9P|1|GA, 4V9P|1|AA, 4V9D|1|CA, 4V9O|1|AA, 4V9O|1|EA, 4WWW|1|RA, 4V9O|1|CA, 4V6C|1|BA, 5J7L|1|CA, 4U27|1|BA, 4V7T|1|BA, 4WF1|1|BA, 5J8A|1|CA, 5JC9|1|CA,

Where? rna.bgsu.edu/rna3dhub/nrlist/release/3.21



Representative Sets of RNA 3D Structures

Release 3.21, 2018-05-11

Resolution threshold

Resolution cutoffs: 1.5A 2.0A 2.5A 3.0A 3.5A 4.0A 20.0A All Download

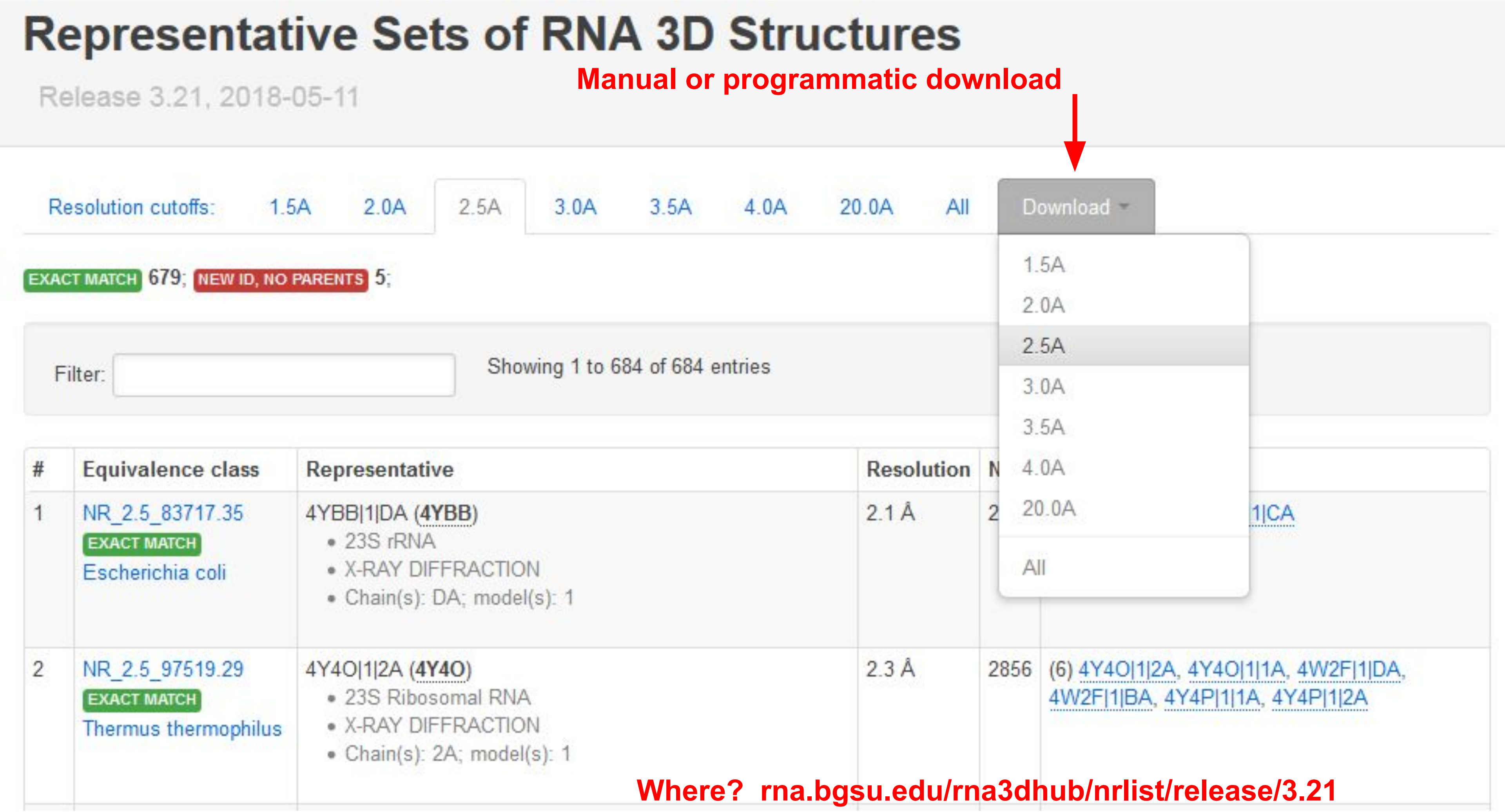
EXACT MATCH 679; NEW ID, NO PARENTS 5;

Filter:

Showing 1 to 684 of 684 entries

#	Equivalence class	Representative	Resolution	Nts	Class members
1	NR_2.5_83717.35 EXACT MATCH Escherichia coli	4YBB|1|DA (**4YBB**) • 23S rRNA • X-RAY DIFFRACTION • Chain(s): DA; model(s): 1	2.1 Å	2883	(2) 4YBB|1|DA, 4YBB|1|CA
2	NR_2.5_97519.29 EXACT MATCH Thermus thermophilus	4Y4O|1|2A (**4Y4O**) • 23S Ribosomal RNA • X-RAY DIFFRACTION • Chain(s): 2A; model(s): 1	2.3 Å	2856	(6) 4Y4O|1|2A, 4Y4O|1|1A, 4W2F|1|DA, 4W2F|1|BA, 4Y4P|1|1A, 4Y4P|1|2A

Where? rna.bgsu.edu/rna3dhub/nrlist/release/3.21



Representative Sets of RNA 3D Structures

Release 3.21, 2018-05-11

Manual or programmatic download

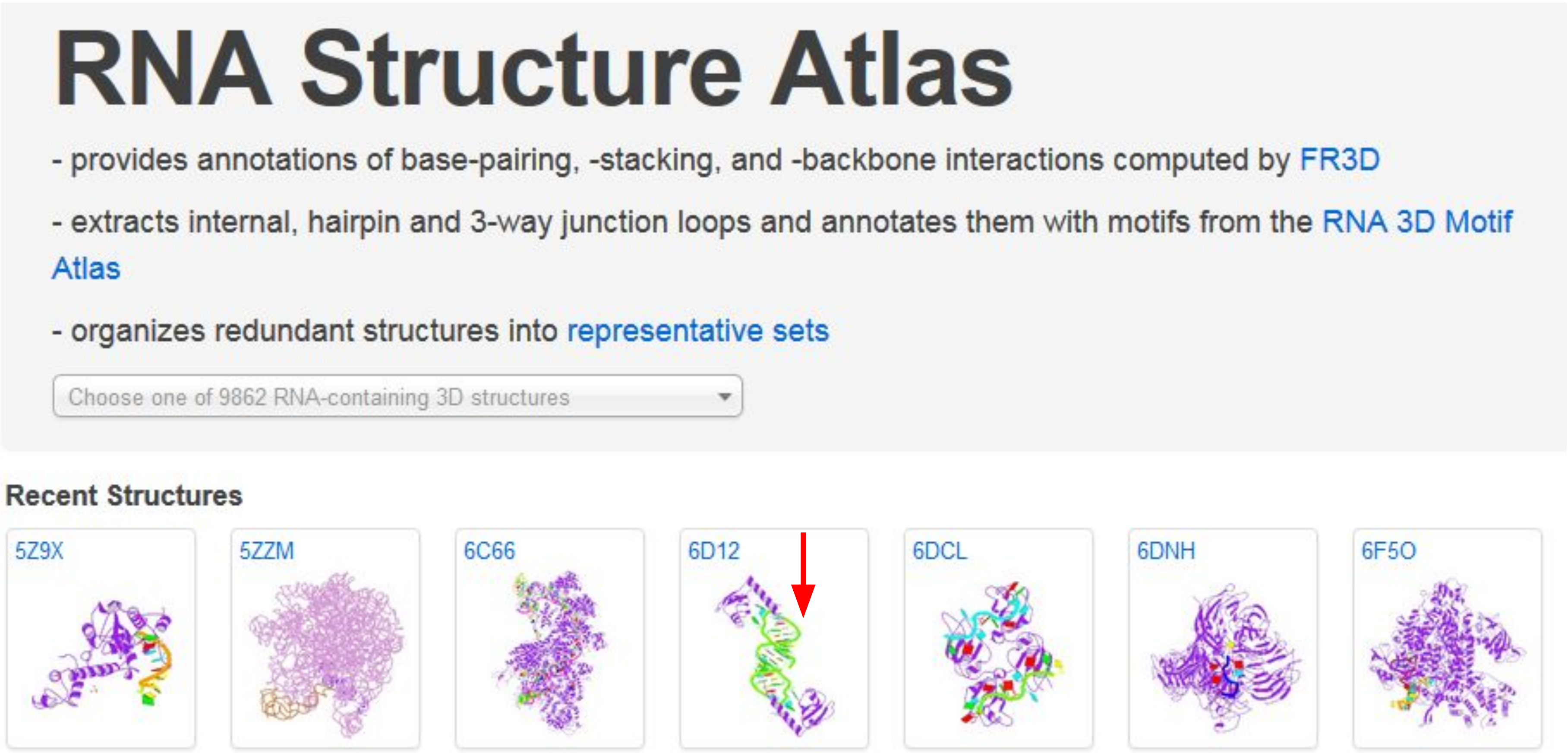
EXACT MATCH 679; NEW ID, NO PARENTS 5;

Showing 1 to 684 of 684 entries

#	Equivalence class	Representative	Resolution	N	
1	NR_2.5_83717.35 EXACT MATCH Escherichia coli	4YBB|1|DA (4YBB) • 23S rRNA • X-RAY DIFFRACTION • Chain(s): DA; model(s): 1	2.1 Å	2	1|CA
2	NR_2.5_97519.29 EXACT MATCH Thermus thermophilus	4Y4O|1|2A (4Y4O) • 23S Ribosomal RNA • X-RAY DIFFRACTION • Chain(s): 2A; model(s): 1	2.3 Å	2856	(6) 4Y4O|1|2A, 4Y4O|1|1A, 4W2F|1|DA, 4W2F|1|BA, 4Y4P|1|1A, 4Y4P|1|2A

Where? rna.bgsu.edu/rna3dhub/nrlist/release/3.21

Annotations of all 3D structures



RNA Structure Atlas

- provides annotations of base-pairing, -stacking, and -backbone interactions computed by FR3D

- extracts internal, hairpin and 3-way junction loops and annotates them with motifs from the RNA 3D Motif Atlas

- organizes redundant structures into representative sets

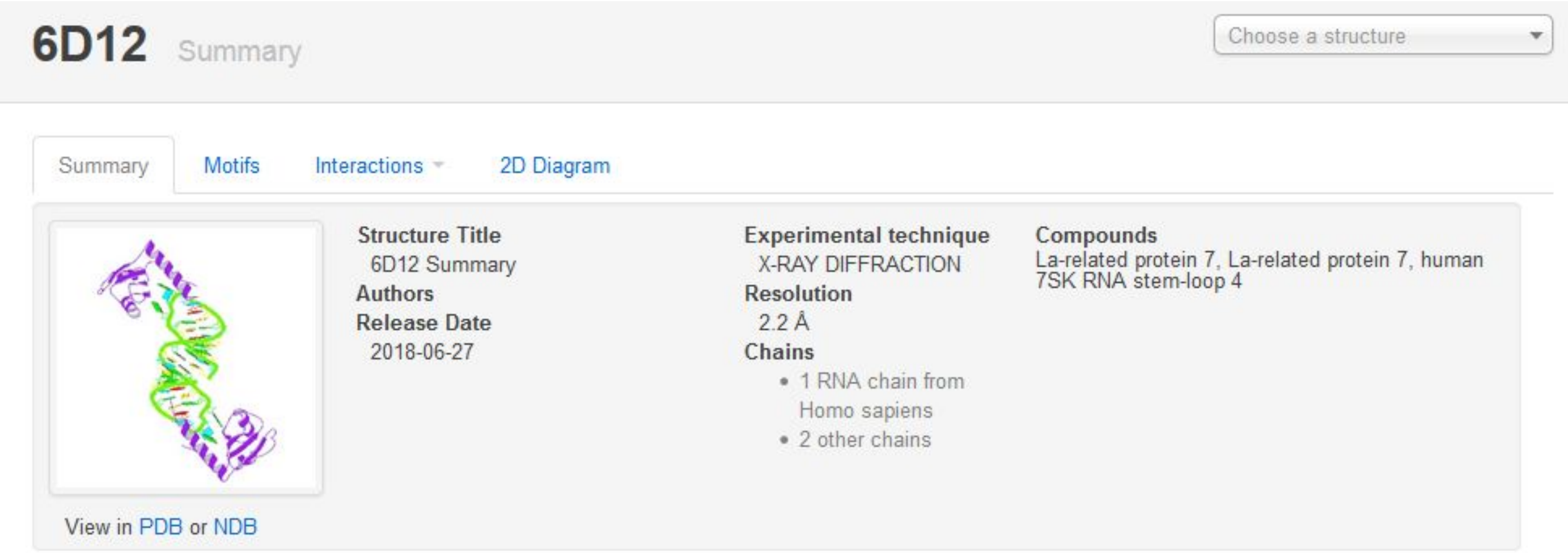
Choose one of 9862 RNA-containing 3D structures ▾

Recent Structures

5Z9X 5ZZM 6C66 6D12 6DCL 6DNH 6F5O

Where? rna.bgsu.edu/rna3dhub/pdb

New structure 6D12



6D12 Summary

Structure Title
6D12 Summary

Authors

Release Date
2018-06-27

Experimental technique
X-RAY DIFFRACTION

Resolution
2.2 Å

Chains
- 1 RNA chain from Homo sapiens
- 2 other chains

Compounds
La-related protein 7, La-related protein 7, human 7SK RNA stem-loop 4

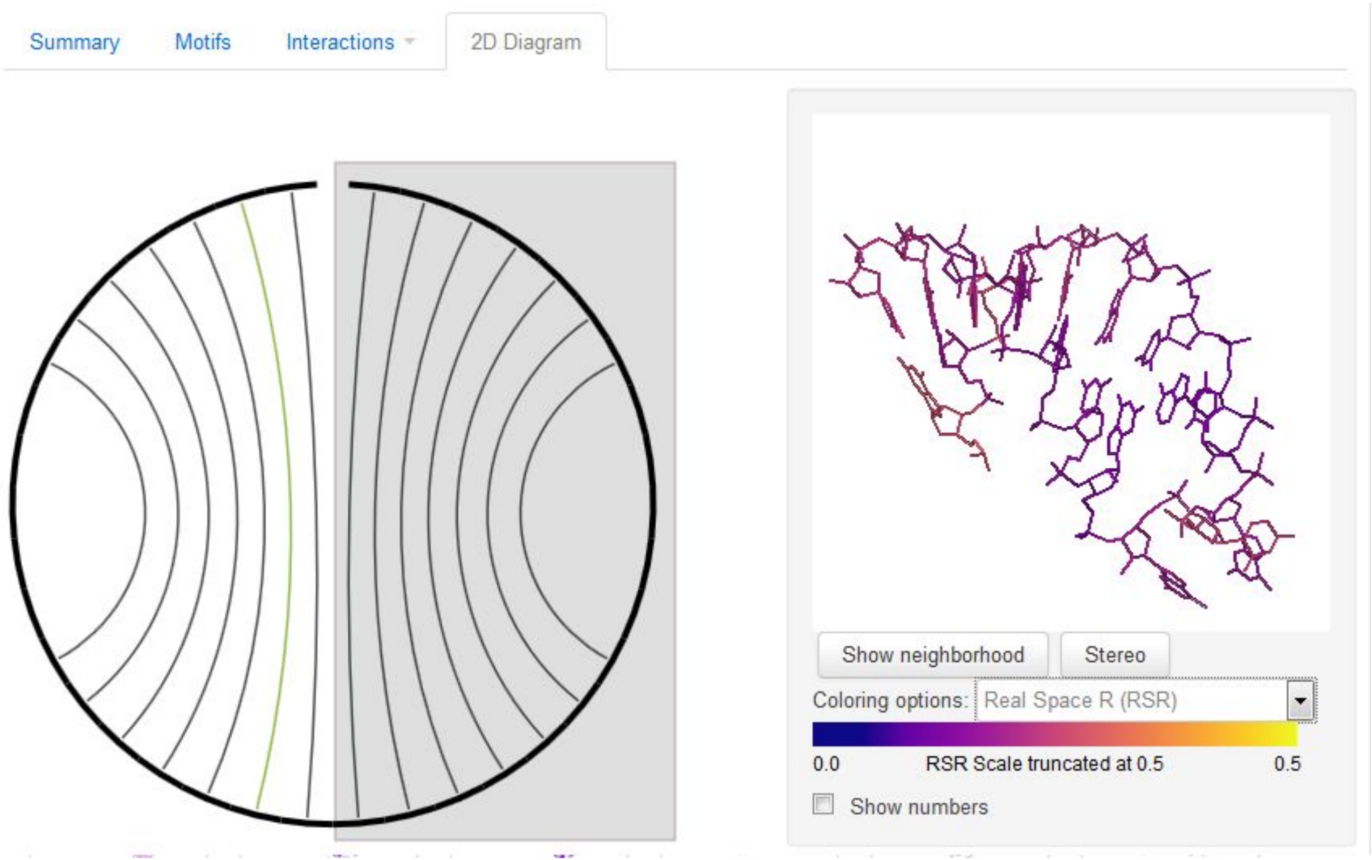
View in PDB or NDB

Where? rna.bgsu.edu/rna3dhub/pdb/6D12

New structure 6D12, circular basepair diagram, RSR

Double helix with one non-WC basepair (green)

Selected nucleotides shown in 3D window, colored by Real Space R (RSR) to indicate structure quality



Coloring options: Real Space R (RSR)

Where? rna.bgsu.edu/rna3dhub/pdb/6D12

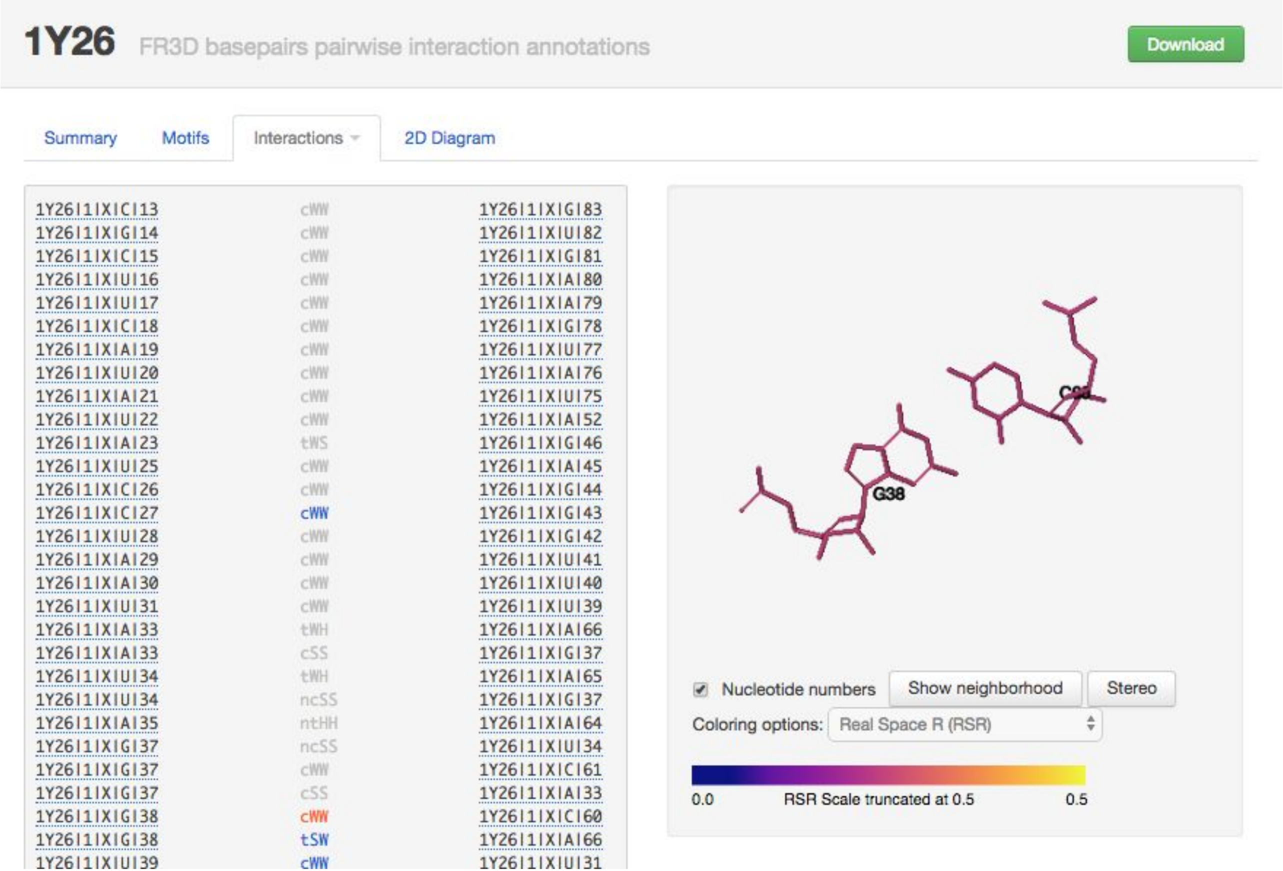
RSR/Z Viewer

We have added a new feature to the 3D Viewer in the BGSU RNA pages that allows users to color residues by RSR and RSRZ values.

The list shows all basepairs in structure 1Y26.

Each basepair can be visualized, colored by RSR value.

Basepair classification limits updated in 2017.



1Y26 FR3D basepairs pairwise interaction annotations

1Y26|1|X|C|13	cWW	1Y26|1|X|G|83
1Y26|1|X|G|14	cWW	1Y26|1|X|U|82
1Y26|1|X|C|15	cWW	1Y26|1|X|G|81
1Y26|1|X|U|16	cWW	1Y26|1|X|A|80
1Y26|1|X|U|17	cWW	1Y26|1|X|A|79
1Y26|1|X|C|18	cWW	1Y26|1|X|G|78
1Y26|1|X|A|19	cWW	1Y26|1|X|U|77
1Y26|1|X|U|20	cWW	1Y26|1|X|A|76
1Y26|1|X|A|21	cWW	1Y26|1|X|U|75
1Y26|1|X|U|22	cWW	1Y26|1|X|A|52
1Y26|1|X|A|23	tWS	1Y26|1|X|G|46
1Y26|1|X|U|25	cWW	1Y26|1|X|A|45
1Y26|1|X|C|26	cWW	1Y26|1|X|G|44
1Y26|1|X|C|27	cWW	1Y26|1|X|G|43
1Y26|1|X|U|28	cWW	1Y26|1|X|G|42
1Y26|1|X|A|29	cWW	1Y26|1|X|U|41
1Y26|1|X|A|30	cWW	1Y26|1|X|U|40
1Y26|1|X|U|31	cWW	1Y26|1|X|U|39
1Y26|1|X|A|33	tWH	1Y26|1|X|A|66
1Y26|1|X|A|33	cSS	1Y26|1|X|G|37
1Y26|1|X|U|34	tWH	1Y26|1|X|A|65
1Y26|1|X|U|34	ncSS	1Y26|1|X|G|37
1Y26|1|X|A|35	ntHH	1Y26|1|X|A|64
1Y26|1|X|G|37	ncSS	1Y26|1|X|U|34
1Y26|1|X|G|37	cWW	1Y26|1|X|C|61
1Y26|1|X|G|37	cSS	1Y26|1|X|A|33
1Y26|1|X|G|38	cWW	1Y26|1|X|C|60
1Y26|1|X|G|38	tSW	1Y26|1|X|A|66
1Y26|1|X|U|39	cWW	1Y26|1|X|U|31

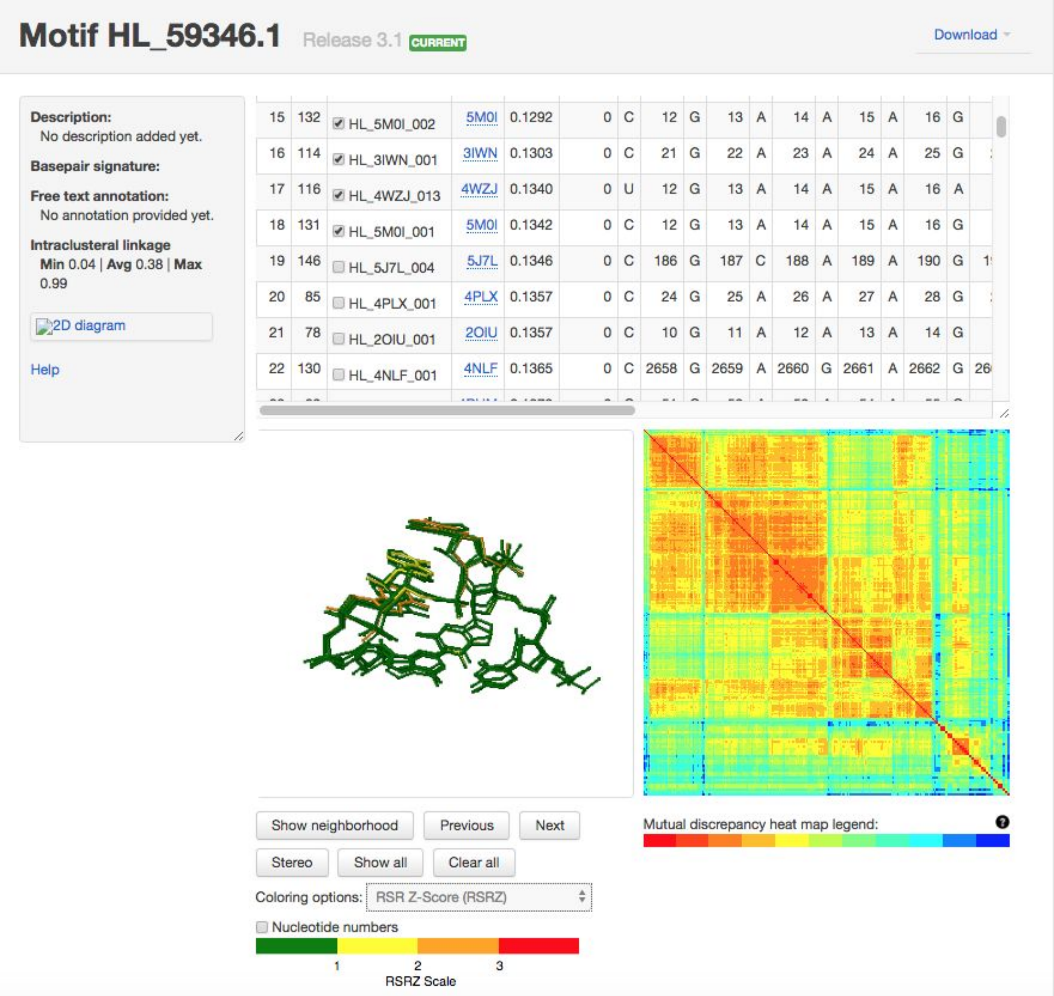
☑ Nucleotide numbers

Coloring options: Real Space R (RSR)

Where? rna.bgsu.edu/rna3dhub/pdb/1Y26

RNA 3D Motif Atlas and RSRZ coloring

Superposed instances of a hairpin loop motif group, colored by RSRZ, using the colorbar shown below the coordinates.



Motif HL_59346.1

Description:
No description added yet.

Basepair signature:

Free text annotation:
No annotation provided yet.

Intraclusteral linkage
Min 0.04 | **Avg** 0.38 | **Max** 0.99

15	132	☑ HL_5M0I_002	5M0I	0.1292	0	C	12	G	13	A	14	A	15	A	16 G
16	114	☑ HL_3IWN_001	3IWN	0.1303	0	C	21	G	22	A	23	A	24	A	25 G
17	116	☑ HL_4WZJ_013	4WZJ	0.1340	0	U	12	G	13	A	14	A	15	A	16 A
18	131	☑ HL_5M0I_001	5M0I	0.1342	0	C	12	G	13	A	14	A	15	A	16 G
19	146	☐ HL_5J7L_004	5J7L	0.1346	0	C	186	G	187	C	188	A	189	A	190 G
20	85	☐ HL_4PLX_001	4PLX	0.1357	0	C	24	G	25	A	26	A	27	A	28 G
21	78	☐ HL_2OIU_001	2OIU	0.1357	0	C	10	G	11	A	12	A	13	A	14 G
22	130	☐ HL_4NLF_001	4NLF	0.1365	0	C	2658	G	2659	A	2660	G	2661	A	2662 G

Coloring options: RSR Z-Score (RSRZ)

☐ Nucleotide numbers

Where? rna.bgsu.edu Motif Atlas, Hairpin

Real Space R (RSR) & RSRZ Server

To support the visualizations, we have built a coordinate server, which returns 3D coordinates of specified residues in PDB format, and an RSR/Z server, which returns RSR and RSRZ values of specified residues.

RSR example, specifying two nucleotides by their Unit ID

http://rna.bgsu.edu/rna3dhub/rest/getRSR?quality=1FJG|1|A|A|16,1FJG|1|A|C|18

RSRZ example, specifying a hairpin loop by its Loop ID

http://rna.bgsu.edu/rna3dhub/rest/getRSRZ?quality=HL_1FJG_001

Applications of the RSR/Z Viewer

Coloring residues by RSR/RSRZ helps users identify interactions or parts of a motif that are not structurally ‘reliable’ or problematic. These are usually regions in a structure that are flexible or involved in local motion. This allows us to determine post-annotation steps that are required to ensure only interactions/motifs that have high structural confidence data are stored in our database.

Motif HL_59346.1 Release 3.1 CURRENT

Description:
No description added yet.

Basepair signature:

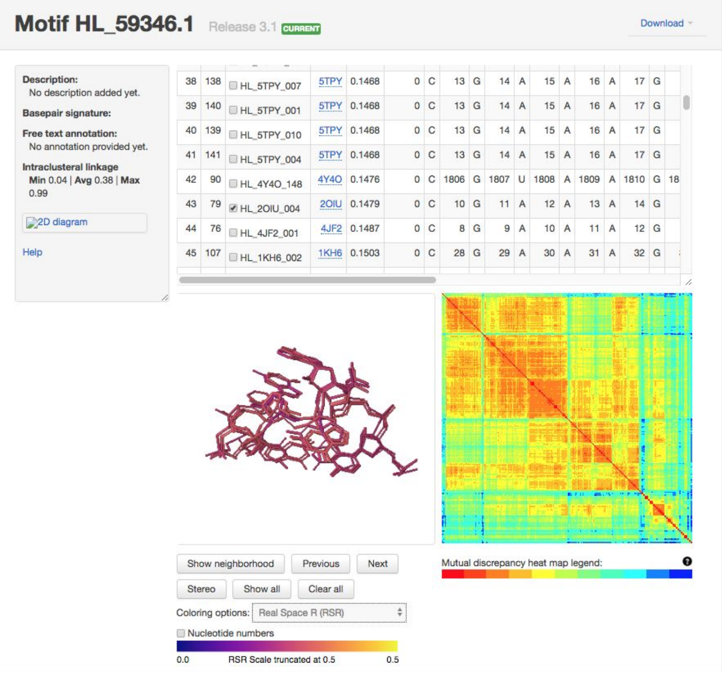
Free text annotation:
No annotation provided yet.

Intraclusteral linkage
Min 0.04 | **Avg** 0.38 | **Max** 0.99

38	138	☐ HL_5TPY_007	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
39	140	☐ HL_5TPY_001	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
40	139	☐ HL_5TPY_010	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
41	141	☐ HL_5TPY_004	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
42	90	☐ HL_4Y4O_148	4Y4O	0.1476	0	C	1806	G	1807	U	1808	A	1809	A	1810	G	18
43	79	☑ HL_2OIU_004	2OIU	0.1479	0	C	10	G	11	A	12	A	13	A	14	G	
44	76	☐ HL_4JF2_001	4JF2	0.1487	0	C	8	G	9	A	10	A	11	A	12	G	
45	107	☐ HL_1KH6_002	1KH6	0.1503	0	C	28	G	29	A	30	A	31	A	32	G	

Coloring options: RSR Z-Score (RSRZ)

☐ Nucleotide numbers



Motif HL_59346.1 Release 3.1 CURRENT

Description:
No description added yet.

Basepair signature:

Free text annotation:
No annotation provided yet.

Intraclusteral linkage
Min 0.04 | **Avg** 0.38 | **Max** 0.99

38	138	☐ HL_5TPY_007	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
39	140	☐ HL_5TPY_001	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
40	139	☐ HL_5TPY_010	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
41	141	☐ HL_5TPY_004	5TPY	0.1468	0	C	13	G	14	A	15	A	16	A	17	G	
42	90	☐ HL_4Y4O_148	4Y4O	0.1476	0	C	1806	G	1807	U	1808	A	1809	A	1810	G	18
43	79	☑ HL_2OIU_004	2OIU	0.1479	0	C	10	G	11	A	12	A	13	A	14	G	
44	76	☐ HL_4JF2_001	4JF2	0.1487	0	C	8	G	9	A	10	A	11	A	12	G	
45	107	☐ HL_1KH6_002	1KH6	0.1503	0	C	28	G	29	A	30	A	31	A	32	G	

Coloring options: Real Space R (RSR)

☐ Nucleotide numbers

Some instances in the GNRA motif group have high average RSR/Z values, and so might need to be filtered out in future releases of the Motif Atlas (HL_2OIU_001, HL_2OIU_004, HL_5F9F_003)

Representative Sets Generation

We present weekly releases (starting from 2011) of RNA structures grouped into sequence/structure equivalence classes on the BGSU RNA Site.

http://rna.bgsu.edu/rna3dhub/nrlist/

Each class includes a representative structure, originally selected as the structure with the most annotated base pairs per nucleotide. Starting with Release 3.0 (structures released through 15 December 2017), all representatives for equivalence classes were selected using Composite Quality Scoring (CQS). We are steadily building the missing releases at the rate of 2-3 per week; the current release is 3.21, structures released through 11 May 2018.