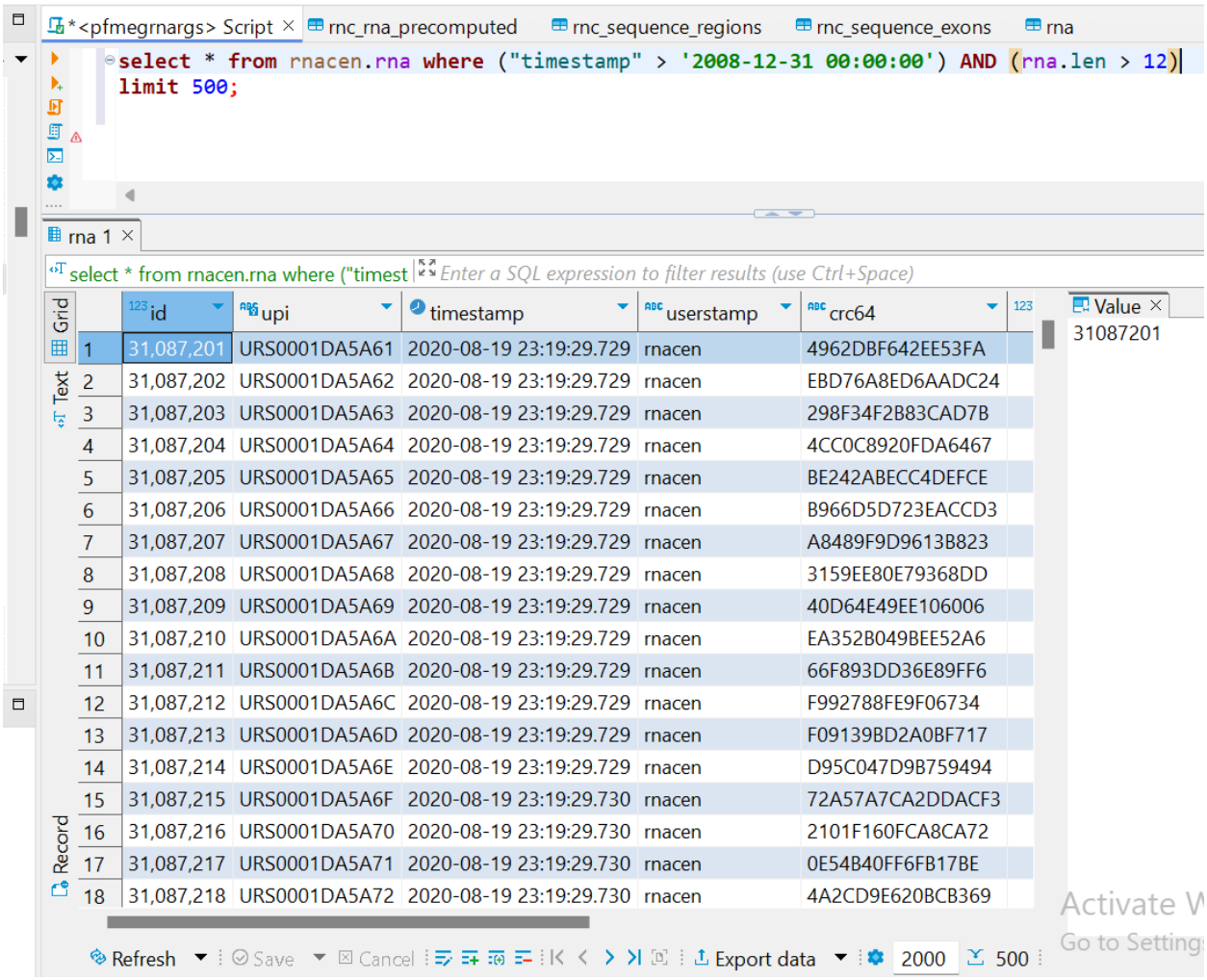
BWT Task-02 Exercise

Submitted By: ABRAR SAEED

Q1

Write a query to get data having length of Rna structures more than 12 with them being added after 2008.



The condition of limit 500 has been added because of large result set achieved after executing query. There are more rows returned instead.

Q2

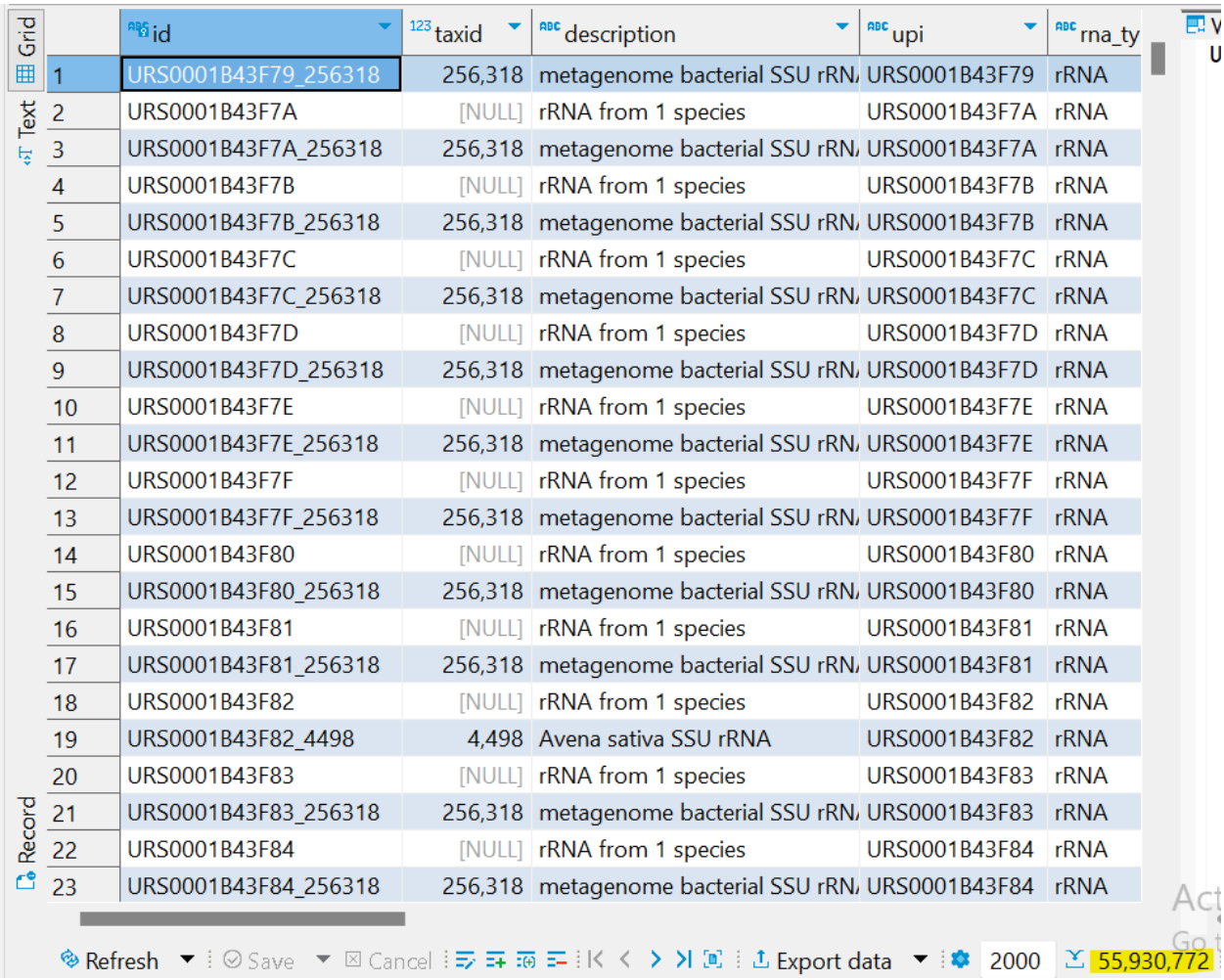
How many pre computed RNA are present that are still active and got their last release update before 2022?

```
SELECT count(distinct rrp.id) FROM rnacen.rnc_rna_precomputed rrp
WHERE rrp.update_date < '2022-01-01' AND (rrp.is_active = true )
```

rnc_rna_precomputed | update_date < '2022-01-01' AND (is_active = true)

	pe	update_date	has_coordinates	databases	is_active
1		2020-08-03	[ ]	SILVA	[v]
2		2020-08-07	[ ]	SILVA	[v]
3		2020-08-03	[ ]	SILVA	[v]
4		2020-08-07	[ ]	SILVA	[v]
5		2020-08-03	[ ]	SILVA	[v]
6		2020-08-07	[ ]	SILVA	[v]
7		2020-08-03	[ ]	SILVA	[v]
8		2020-08-07	[ ]	SILVA	[v]
9		2020-08-03	[ ]	SILVA	[v]
10		2020-08-07	[ ]	SILVA	[v]
11		2020-08-03	[ ]	SILVA	[v]
12		2020-08-07	[ ]	SILVA	[v]
13		2020-08-03	[ ]	SILVA	[v]
14		2020-08-07	[ ]	SILVA	[v]
15		2020-08-03	[ ]	SILVA	[v]
16		2020-08-07	[ ]	SILVA	[v]
17		2020-08-03	[ ]	SILVA	[v]
18		2020-08-07	[ ]	SILVA	[v]
19		2020-08-03	[ ]	SILVA	[v]
20		2020-08-07	[ ]	SILVA	[v]
21		2020-08-03	[ ]	SILVA	[v]
22		2020-08-07	[ ]	SILVA	[v]
23		2020-08-03	[ ]	SILVA	[v]

Refresh | Save | Cancel | Export data | 2000 | 55,930,772

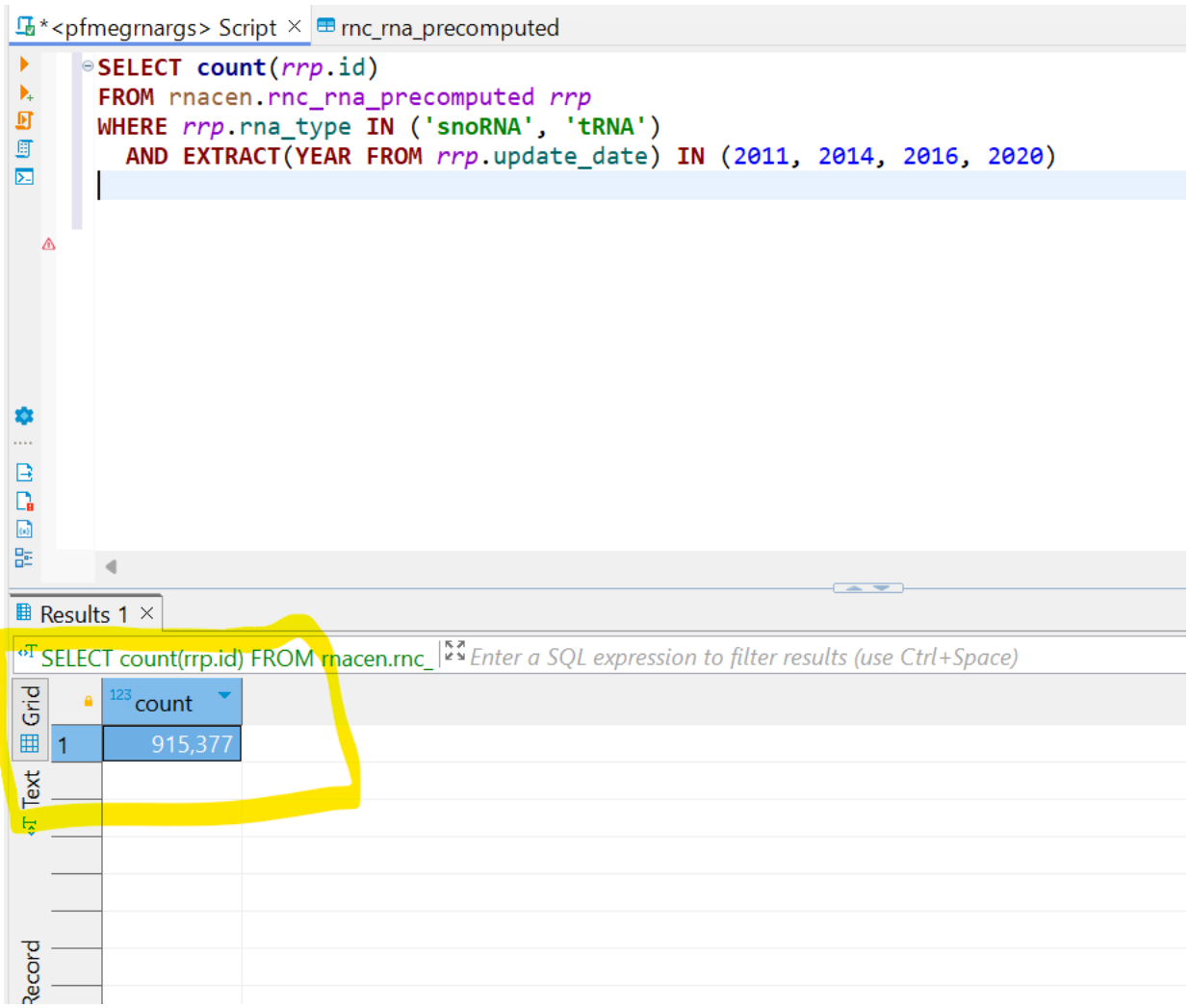


	id	taxid	description	upi	rna_ty
1	URS0001B43F79_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F79	rRNA
2	URS0001B43F7A	[NULL]	rRNA from 1 species	URS0001B43F7A	rRNA
3	URS0001B43F7A_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7A	rRNA
4	URS0001B43F7B	[NULL]	rRNA from 1 species	URS0001B43F7B	rRNA
5	URS0001B43F7B_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7B	rRNA
6	URS0001B43F7C	[NULL]	rRNA from 1 species	URS0001B43F7C	rRNA
7	URS0001B43F7C_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7C	rRNA
8	URS0001B43F7D	[NULL]	rRNA from 1 species	URS0001B43F7D	rRNA
9	URS0001B43F7D_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7D	rRNA
10	URS0001B43F7E	[NULL]	rRNA from 1 species	URS0001B43F7E	rRNA
11	URS0001B43F7E_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7E	rRNA
12	URS0001B43F7F	[NULL]	rRNA from 1 species	URS0001B43F7F	rRNA
13	URS0001B43F7F_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F7F	rRNA
14	URS0001B43F80	[NULL]	rRNA from 1 species	URS0001B43F80	rRNA
15	URS0001B43F80_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F80	rRNA
16	URS0001B43F81	[NULL]	rRNA from 1 species	URS0001B43F81	rRNA
17	URS0001B43F81_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F81	rRNA
18	URS0001B43F82	[NULL]	rRNA from 1 species	URS0001B43F82	rRNA
19	URS0001B43F82_4498	4,498	Avena sativa SSU rRNA	URS0001B43F82	rRNA
20	URS0001B43F83	[NULL]	rRNA from 1 species	URS0001B43F83	rRNA
21	URS0001B43F83_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F83	rRNA
22	URS0001B43F84	[NULL]	rRNA from 1 species	URS0001B43F84	rRNA
23	URS0001B43F84_256318	256,318	metagenome bacterial SSU rRN	URS0001B43F84	rRNA

The highlighted number at the bottom right corner (55,930,772) tells the number of rows returned based on the conditions of query that are all those pre-computed RNA that are active and have last update before 2022.

Q3

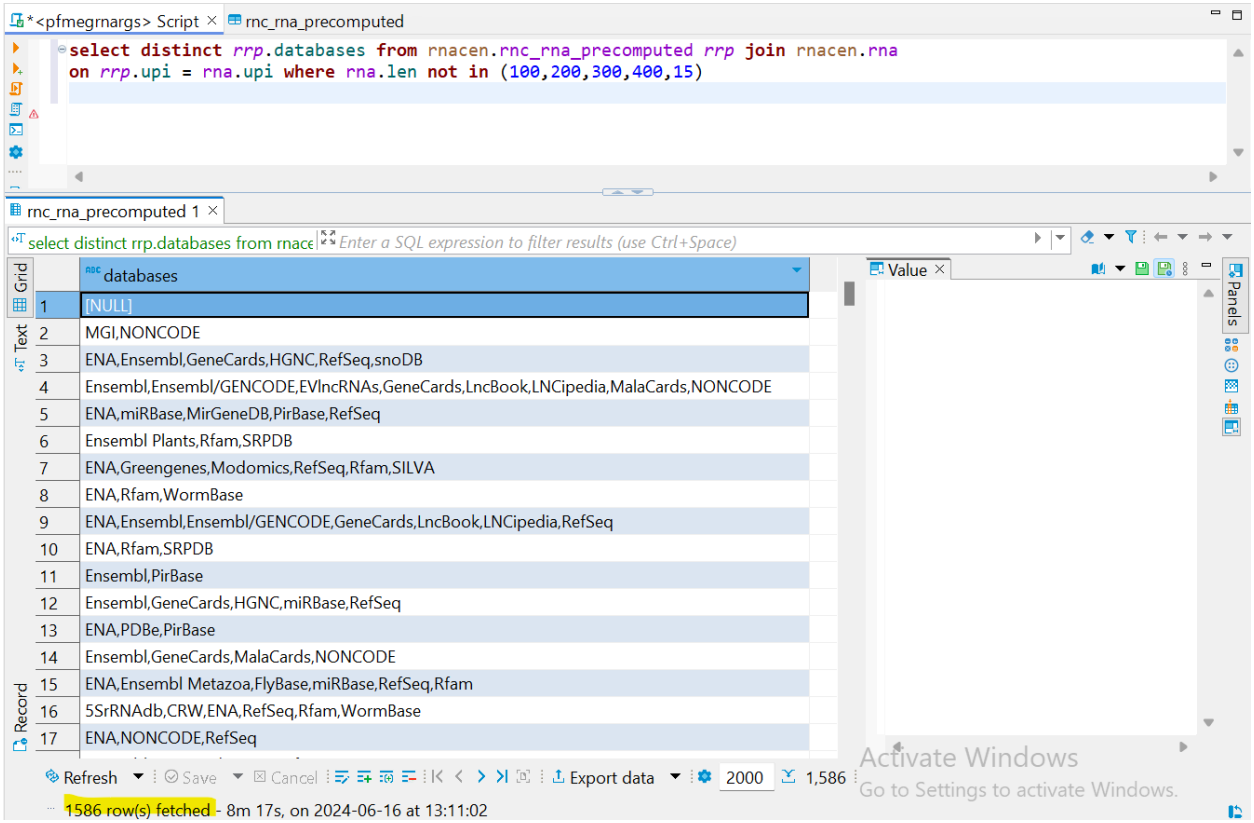
How many total pre computed RNA records for snoRNA and tRNA were recorded in 2011, 2016, 2014, and 2020?

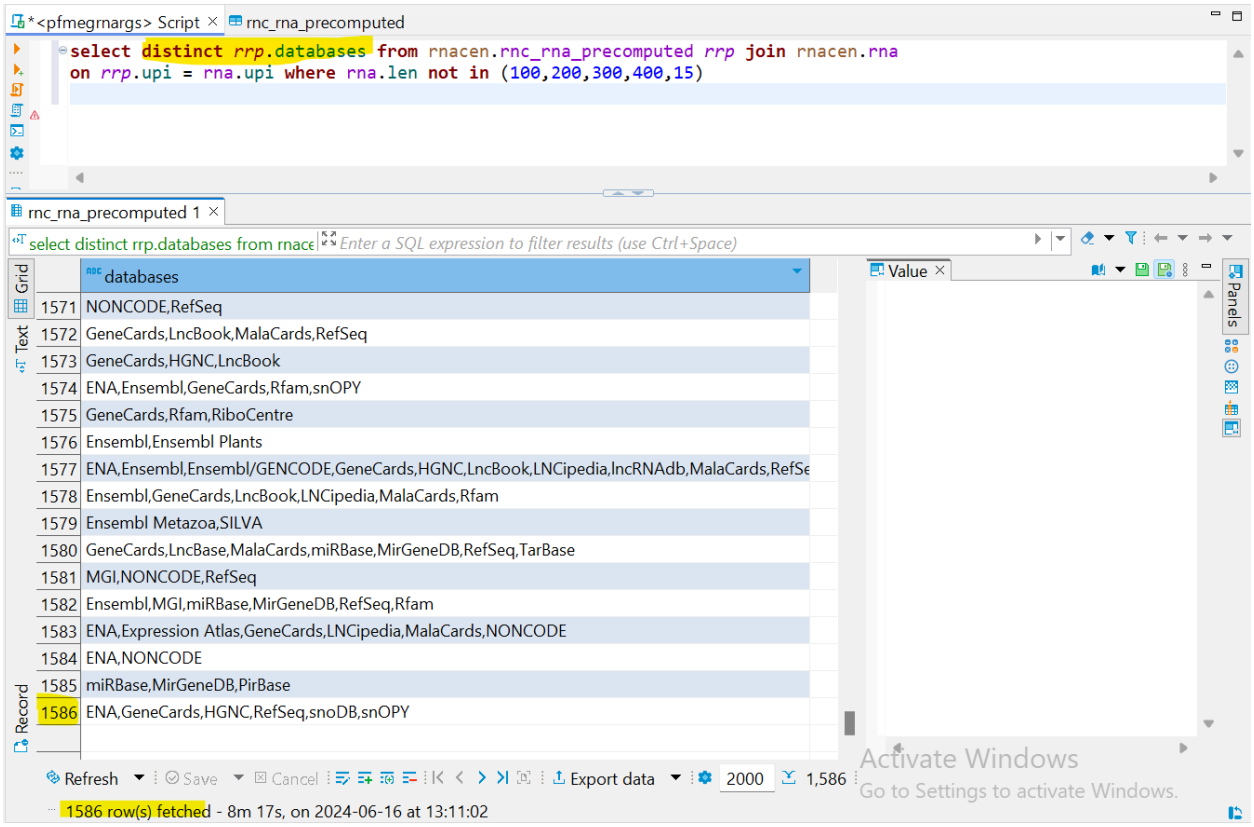


Q4

Can you give me the names of all databases built for RNA with minimum length other than 100, 200, 300, 400, and 15?

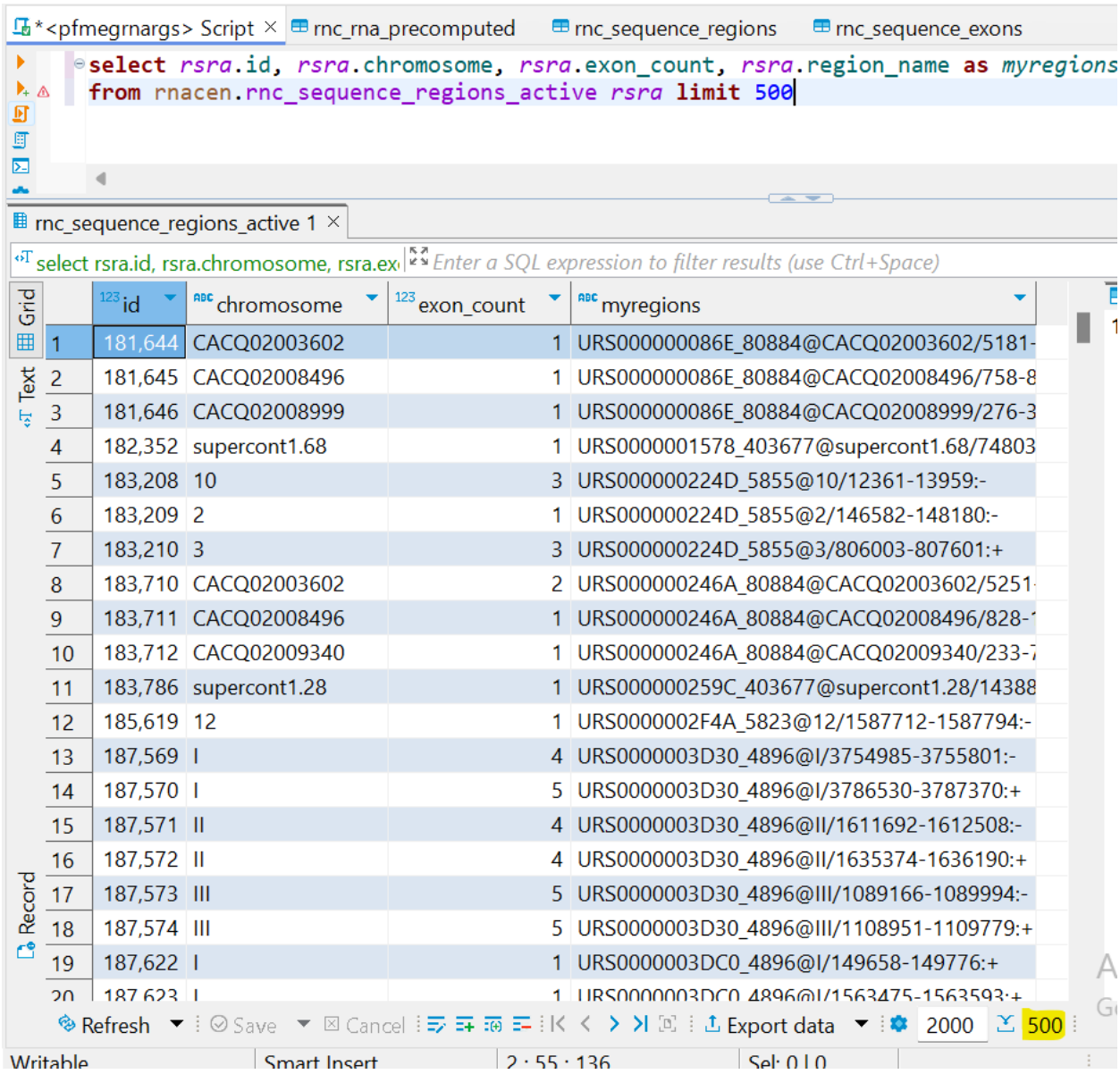
```
select distinct rrp.databases from rnacen.rnc_rna_precomputed rrp join rnacen.rna
on rrp.upi = rna.upi where rna.len not in (100,200,300,400,15)
```




*<pfmegrnargs> Script
rnc_rna_precomputed
select distinct rrp.databases from rnacen.rnc_rna_precomputed rrp join rnacen.rna
on rrp.upi = rna.upi where rna.len not in (100,200,300,400,15)
rnc_rna_precomputed 1
select distinct rrp.databases from rnace
Enter a SQL expression to filter results (use Ctrl+Space)
Grid
Text
Record
databases
1571 NONCODE,RefSeq
1572 GeneCards,LncBook,MalaCards,RefSeq
1573 GeneCards,HGNC,LncBook
1574 ENA,Ensembl,GeneCards,Rfam,snOPY
1575 GeneCards,Rfam,RiboCentre
1576 Ensembl,Ensembl Plants
1577 ENA,Ensembl,Ensembl/GENCODE,GeneCards,HGNC,LncBook,LNCipedia,lncRNAdb,MalaCards,RefSe
1578 Ensembl,GeneCards,LncBook,LNCipedia,MalaCards,Rfam
1579 Ensembl Metazoa,SILVA
1580 GeneCards,LncBase,MalaCards,miRBase,MirGeneDB,RefSeq,TarBase
1581 MGI,NONCODE,RefSeq
1582 Ensembl,MGI,miRBase,MirGeneDB,RefSeq,Rfam
1583 ENA,Expression Atlas,GeneCards,LNCipedia,MalaCards,NONCODE
1584 ENA,NONCODE
1585 miRBase,MirGeneDB,PirBase
1586 ENA,GeneCards,HGNC,RefSeq,snoDB,snOPY
Value
Panels
Refresh
Save
Cancel
Export data
2000
1,586
1586 row(s) fetched - 8m 17s, on 2024-06-16 at 13:11:02
Activate Windows
Go to Settings to activate Windows.

Q5

Can you get complete 500 records of sequences for active regions and name your column as myregions in which you are getting the region name column value. Then tell me what different chromosomes with exon_count we have for regions including center, east and north using the name you set for your column?



```sql
select
    rsra.region_name AS myregions,
    array_agg(DISTINCT rsra.chromosome) AS chromosomes,
    array_agg(DISTINCT rsra.exon_count) AS exon_counts
from
    rnacen.rnc_sequence_regions_active rsra
group by
    rsra.region_name
limit 500;
```

rnc_sequence_regions_active 1

select rsra.region_name AS myregions, *Enter a SQL expression to filter results (use Ctrl+Space)*

Grid | Text | Record

	myregions	chromosomes	exon_counts
1	URS0000000019_10029@JH000057.1/2028144-20281	{JH000057.1}	{1}
2	URS000000001B_7091@scaf189/42519-42546:+	{scaf189}	{1}
3	URS000000003E_10116@8/68353160-68353189:-	{8}	{1}
4	URS0000000055_9606@6/57171004-57171098,57173	{6}	{3}
5	URS0000000055_9606@6/57171005-57171098,57173	{6}	{3}
6	URS0000000056_10116@8/54611676-54611705:-	{8}	{1}
7	URS0000000064_10116@8/68354471-68354499:-	{8}	{1}
8	URS0000000074_10090@17/27568298-27568324:+	{17}	{1}
9	URS0000000081_3702@1/11880517-11880540:-	{1}	{1}
10	URS0000000085_46245@4_group3/272327-272405:+	{4_group3}	{1}
11	URS0000000085_7234@scaffold_1/1745491-1745569	{scaffold_1}	{1}
12	URS0000000088_10090@12/98388598-98388627:+	{12}	{1}
13	URS000000008C_10090@16/10995055-10995084:+	{16}	{1}
14	URS0000000096_9606@15/51291230-51291259:-	{15}	{1}
15	URS0000000098_9606@10/17403508-17403618:+	{10}	{1}

Refresh | Save | Cancel | Export data | 2000 | 500