The following Query is a report query that returns all of the "Kozaks" for a given Species. That is the ten characters before and after the start index for a given ORF in its strain's genome. This is used to generate the consensus sequence.

SELECT UPPER(SUBSTRING(strains.genome, orf.startIndex - 10, 20)) AS Kozak FROM OpenReadingFrames orf

JOIN Strains AS strains ON orf.strainID = strains.strainID

where orf.strainID in (Select strainID FROM Strains s

JOIN Species AS species ON s.specID = species.specID

WHERE sName = 'Avian Coronavirus')

## Kozak

CAGTGACAACATGGCTTCAA CTGGTAAGAGATGTTGGTTC AGTCTGTTTAATGATTCAAA GTGTAGACTAATGTTAGATT AGTTCAGGTGATGGTGAACT ACTCCAGCAAATGTCCATGA AATACAGACGATGAAATGGC AATCGCTGGTATGAATAATA ATCTTTTATCATGGCAAGTG CAGTGACAACATGGCTTCAA CTGGTAAGAGATGTTGGCTC AGTCCGTTTAATGATTCAAA GTCTAGACTAATGTTAGATT AGTTCAGGTGATGGTGAACT AATCCAGCAAATGTCCACGA AATACGGACGATGAAATGGC AATCGCTGGTATGAATAATA GTCTTTTGTCATGGCAAGCG CAGTGACAACATGGCTTCAA CTGGTAAGAGATGTTGGTAA AGTCTGTTTAATGATCCAAA GTCTAGACTAATGTTAAACT TGTTCAGGTGATGATGAATT AATTTTCAAGATGTCCAACG AATACAGACGATGAAATGGC 2. The following Query is a report query that returns all of the ORFs for a given Species (that includes all of its Strains) that work with our consensus algorithm. That is, only the ORFs that do not have any characters other than A, or if it has an A, any other characters are fine. For example, 2B is not okay, but 2AB is okay.

SELECT DISTINCT orfID FROM OpenReadingFrames where strainID in

```
(SELECT strainID FROM Strains s join Species as species on s.specID =
species.specID where sName = "Avian Coronavirus" and (orfID like '%a%' or
orfID not regexp '[b-z]')
)
```

orfID
1ab
2
3a
4
5
6a
7

3. The following is a Query that lists all of the Strains and its ORFs for a given Species name. This query without the where statement is also useful for seeing all of the Strains and their ORFs in the database.

SELECT species.sName, strains.strainID, orf.orfID FROM

Species species

JOIN Strains AS strains ON strains.specID = species.specID

JOIN OpenReadingFrames AS orf ON orf.strainID = strains.strainID

WHERE species.sName = "Avian Coronavirus"

ORDER BY sName, strainID, orfID asc

sName 🔺 1	strainID 🔺 2	orfID △ 3
Avian Coronavirus	KX258195	1ab
Avian Coronavirus	KX258195	2
Avian Coronavirus	KX258195	3a
Avian Coronavirus	KX258195	3b
Avian Coronavirus	KX258195	4
Avian Coronavirus	KX258195	5
Avian Coronavirus	KX258195	6a
Avian Coronavirus	KX258195	6b
Avian Coronavirus	KX258195	7
Avian Coronavirus	KY626044	1ab
Avian Coronavirus	KY626044	2
Avian Coronavirus	KY626044	3a
Avian Coronavirus	KY626044	3b
Avian Coronavirus	KY626044	4
Avian Coronavirus	KY626044	5
Avian Coronavirus	KY626044	6a
Avian Coronavirus	KY626044	6b
Avian Coronavirus	KY626044	7
Avian Coronavirus	KY626045	1ab
Avian Coronavirus	KY626045	2
Avian Coronavirus	KY626045	3a
Avian Coronavirus	KY626045	3b
Avian Coronavirus	KY626045	4
Avian Coronavirus	KY626045	5
Avian Coronavirus	KY626045	6a
Avian Coronavirus	KY626045	6b
Avian Coronavirus	KY626045	7

4. The following query is a report query that lists all of the publications and their publisher as well as the year it was published.

SELECT publishers.name AS Publisher, publications.title AS Title, publications.pYear AS Year FROM Publishers publishers

JOIN Publisher\_Publication AS pp ON
publishers.publisherID = pp.publisherID
JOIN Publications AS publications ON
publications.pubID = pp.pubID
ORDER BY Publisher ASC

Publisher 🔺 1	Title	Year
CIS560	CIS560Publication	2017
GenBank	Genome Sequences of Simian Hemorrhagic Fever Virus	2014
GenBank	Genomic RNA sequence of canine coronavirus strain	2009
GenBank	Genomic, phylogenetic, and recombinational analysi	2012
GenBank	Full-genome sequence of pantropic canine coronavir	2015
GenBank	Characterization of a recombinant canine coronavir	2012
GenBank	Feline infectious peritonitis virus undergoes a lo	2013
GenBank	Molecular characterization of feline coronavirus i	2017
GenBank	Feline Coronavirus ORF classification	2009
GenBank	Genomic Sequencing of three Strains of Porcine Rep	2014
GenBank	Genetic diversity of porcine reproductive and resp	2016
GenBank	Genetic variant recombination analysis from PCR am	2014
GenBank	Isolation, propagation, genome analysis and epidem	2014
GenBank	First complete genome of a Brazil type Avian coron	2016
GenBank	Illumina MiSeq sequence of 15 Minnesota PRRSV stra	2015
Infectious Diseases, Animal Sciences Group, Edelhe	Two single amino acid changes in GP2a are responsi	2004
Journal of General Virology	North American and European porcine reproductive a	1999
Journal of General Virology	Characterization of a recombinant canine coronavir	2012
Journal of General Virology	Human Respiratory Coronavirus OC43: Genetic Stabil	2004
Journal of General Virology	Comparative Analysis of 22 Coronavirus HKU1 Genome	2006
Journal of General Virology	Comparative analysis of complete genome sequences	2008
Science Direct	Sequence of the genome of lactate dehydrogenase-el	1995
Science Direct	Complete genomic sequence and phylogenetic analysi	1993
Science Direct	Analysis of simian hemorrhagic fever virus (SHFV)	1995
Science Direct	Molecular characterization of the 3' terminus of t	1995

5. The following Query lists all of the proteins and their associated ORF and Strain that are referenced by multiple ORFs.

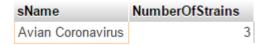
SELECT strainID, orfID, pID FROM OpenReadingFrames

WHERE pID in (SELECT pID from OpenReadingFrames group by pID having count(\*) > 1)

←7	Г→		$\overline{}$	strainID	orflD	pID
		<b>≩</b> сору	Delete	AF046869	1a	0
		<b>≩</b> сору	Delete	AY588319	1a	0
		<b>≩</b> сору	Delete	HQ315836	1a	0
		<b>≩</b> сору	Delete	HQ315837	1a	0
		<b>≩</b> сору	Delete	KM371108	1a	0
		<b>≩</b> сору	Delete	KM371109	1a	0
		<b>≩</b> сору	Delete	KM371110	1a	0
		<b>≩</b> сору	Delete	KM371111	1a	0
		<b>≩</b> сору	Delete	KM373784	1a	0
		<b>≩</b> сору	Delete	KP283403	1a	0
		<b>≩</b> сору	Delete	KP283410	1a	0
		<b>≩</b> сору	Delete	KP283411	1a	0
		<b>≩</b> сору	Delete	KP283413	1a	0
		<b>≩</b> сору	Delete	KP283416	1a	0
		<b>≩</b> сору	Delete	KT257944	1a	0
		<b>≩</b> сору	Delete	KT257950	1a	0
		<b>≩</b> сору	Delete	KT257953	1a	0
		<b>≩</b> сору	Delete	L13298	1a	0
		<b>≩</b> сору	Delete	LDU15146	1a	0
		<b>≩</b> сору	Delete	NC_001639	1a	0
		<b>≩</b> сору	Delete	NC_003092	1a	0
	Edit	<b>≩≟</b> Copy	Delete	AF046869	1b	1
	Edit	<b>≩-</b> Copy	Delete	AY588319	1b	1
	Edit	<b>≟</b> Copy	Delete	HQ315836	1b	1
	Ø Edit	<b>≩</b> € Copy	Delete	HQ315837	<b>1</b> b	1

6. The following Query is a question query with the question: How many Strains are there for the given Species?

SELECT species.sName, count(\*) as NumberOfStrains FROM Strains as strains
JOIN Species AS species ON
species.specID = strains.specID
WHERE species.sName = "Avian Coronavirus"



7. The following Query is a question query with the question: What researchers worked on the given publication?

SELECT publications.title AS Publication, researchers.rName AS Name FROM Publications AS publications

JOIN Publication\_Researcher AS pr ON pr.pubID = publications.pubID

JOIN Researchers AS researchers ON researchers.rID = pr.rID

WHERE publications.title = "Genomic RNA sequence of canine coronavirus strain"

Publication	Name
Genomic RNA sequence of canine coronavirus strain	Ling-Ling Chueh
Genomic RNA sequence of canine coronavirus strain	Bi-Ling Su

8. The following Query is a question query with the question:

What are the ORFs associated with the given Strain?

Select orfID AS ORFS FROM OpenReadingFrames WHERE strainID = "AF046869"

ORFS
1a
1b
2
3
4
5
6
7

9. The following Query is a question query with the question: Which publications were published before 2000?

SELECT title AS Title, pYear as Year FROM Publications WHERE pYear < 2000

Title	Year
North American and European porcine reproductive a	
Sequence of the genome of lactate dehydrogenase-el	
Complete genomic sequence and phylogenetic analysi	
Analysis of simian hemorrhagic fever virus (SHFV)	
Molecular characterization of the 3' terminus of t	

10. The following Query is a question query with the question: What are the Publishers who have "Journal" in their name (finding journals):

SELECT name AS Name FROM Publishers WHERE name like '%Journal%'

## Name

Journal of General Virology

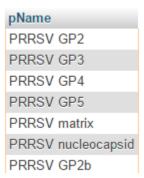
11. The following Query is a question query with the question: What Researchers are from Kansas State University?

SELECT rName AS Name FROM Researchers WHERE rOrg like '%Kansas State%'



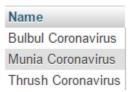
12. The following Query is a question query with the question: Which proteins are PRRSV proteins? (Have that in their name)

SELECT pName FROM Proteins WHERE pName LIKE '%PRRSV%'



13. The following Query is a question query with the question: What Species are a part of the Genus "Deltacoronavirus"?

SELECT sName AS Name FROM Species WHERE sGenus = "Deltacoronavirus"



14. The following Query is a question query with the question: Which proteins are non-structural?

SELECT pName AS Name FROM Proteins AS proteins

JOIN NonStructuralProteins AS nsp ON

nsp.pID = proteins.pID

Name
Arterivirus replicase polyprotein
Arterivirus replicase complex
Coronavirus replicase complex
CCoV ORF3a NSP
CCoV ORF3b NSP
CCoV ORF3c NSP
CCoV ORF7a NSP
CCoV ORF7b NSP
FCoV ORF3a NSP
FCoV ORF3b NSP
FCoV ORF3c NSP
FCoV ORF7a NSP
FCoV ORF7b NSP
HCoV ORF2a NSP
HCoV ORF4 NSP
BuCoV ORF6 NSP
BuCoV ORF8 NSpP
MuCoV ORF6 NSP
MuCoV ORF8 NSpP
TCoV ORF6 NSP
TCoV ORF8 NSpP
ACoV ORF3a NSP
ACoV ORF3b NSP
ACoC ORF6a NSP
ACoV ORF6b NSP

15. The following query is a question query with the question: What publications reference the given strain?

SELECT sp.strainID AS Strain, publications.title AS Title FROM
Strain\_Publication AS sp

JOIN Publications AS publications ON
publications.pubID = sp.pubID

WHERE strainID = "NC\_003092"

Strain	Title
NC_003092	Analysis of simian hemorrhagic fever virus (SHFV)
NC_003092	Molecular characterization of the 3' terminus of t