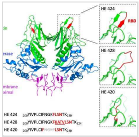


Version 3

Sep 22, 2022

Phylogenetic Analysis of Complete Bovine Coronavirus Genome Sequences V.3

Aspen M Workman[1], Tara G. McDaneld[1], Gregory P Harhay[1], Subha Das[2], John Dustin Loy[3], Benjamin M. Hause[2]

[1]United States Department of Agriculture (USDA) Agricultural Research Service (ARS), US Meat Animal Research Center (USM ARC), State Spur 18D, Clay Center, NE 68933, USA;

[2]South Dakota State University, Brookings, SD 57007, USA;

[3]Nebraska Veterinary Diagnostic Center, School of Veterinary Medicine and Biomedical Sciences, University of Nebraska-Lincoln, 4040 East Campus Loop N, Lincoln, NE 68503-0907

1 Works for me Share

dx.doi.org/10.17504/protocols.io.kqdg3pyeql25/v3

Gregory P Harhay
United States Department of Agriculture (USDA) Agricultural ...

DISCLAIMER

•The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.
•Mention of trade names or commercial products in this presentation is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.
•USDA is an equal opportunity provider and employer

ABSTRACT

Bovine coronavirus (BCoV) has spilled over to many species, including humans, where the host range variant coronavirus OC43 is endemic. Balance of the opposing activities of the surface spike (S) and hemagglutinin esterase (HE) glycoproteins control virion avidity which is critical for interspecies transmission and host adaptation. Here, 78 genomes were sequenced directly from clinical samples collected between 2013 and 2022 from cattle in 12 states, primarily in the Midwestern U.S. Relatively little genetic diversity was observed, with genomes having >98% nucleotide identity. Eleven isolates collected between 2020 and 2022 from four states (Nebraska, Colorado, California, and Wisconsin) contained a 12-nucleotide insertion in the receptor-binding domain (RBD) of the HE gene identical to one recently reported in China, and a single genome from Nebraska collected in 2020 contained a novel 12-nucleotide deletion in the HE gene RBD. Isogenic HE proteins containing either the insertion or deletion in the HE RBD maintained esterase activity and the ability to bind bovine submaxillary mucin, a substrate enriched in the receptor 9-*O*-acetylated-sialic acid, despite modeling that predicted structural changes in the HE R3 loop critical for receptor binding. The emergence of BCoV with structural variants in the RBD raises the possibility of further interspecies transmission.

DOI

dx.doi.org/10.17504/protocols.io.kqdg3pyeql25/v3

PROTOCOL CITATION

Aspen M Workman, Tara G. McDaneld, Gregory P Harhay, Subha Das, John Dustin Loy, Benjamin M. Hause 2022. Phylogenetic Analysis of Complete Bovine Coronavirus Genome Sequences. **protocols.io**
https://protocols.io/view/phylogenetic-analysis-of-complete-bovine-coronavir-cgxntxme
Version created by Gregory P Harhay

FUNDERS ACKNOWLEDGEMENT

National Institute of Food and Agriculture
Grant ID: 2019-67015-29847

USDA ARS CRIS
Grant ID: 3040-32000-036-00D

National Institutes of Allergy and Infectious Diseases
Grant ID: 1R21AI162594-01

MANUSCRIPT CITATION please remember to cite the following publication along with this protocol

Recent emergence of bovine coronavirus variants with mutations in the hemagglutinin-esterase receptor binding domain in U.S. cattle (submitted)

KEYWORDS

coronavirus, hemagglutinin-esterase, spike protein, bovine respiratory disease, calf diarrhea, variant

CREATED

Sep 21, 2022

LAST MODIFIED

Sep 22, 2022

PROTOCOL INTEGER ID

70350

DISCLAIMER:

•The findings and conclusions in this presentation are those of the authors and should not be construed to represent any official USDA or U.S. Government determination or policy.
•Mention of trade names or commercial products in this presentation is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA.
•USDA is an equal opportunity provider and employer

Software

1

MAFFT 7.450
by Kazutaka Katoh

2

RAxML 8.2.12
UBUNTU Linux ` 20.4
source by Alexandros Stamatakis

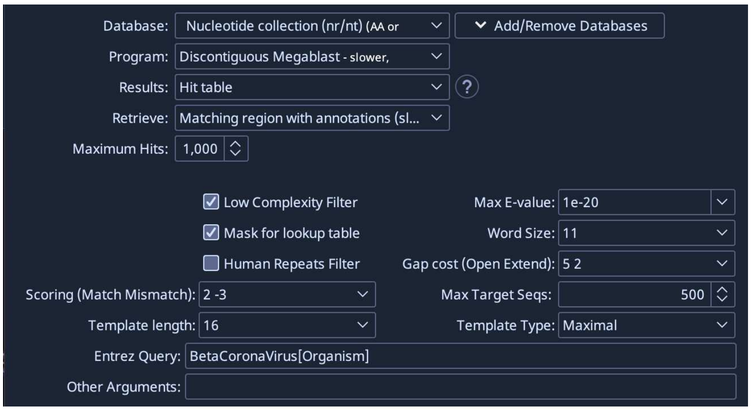
3

FAST --- Fast Analysis of Sequences Toolbox 1.7
UBUNTU Linux 22 LTS
source by Travis Lawrence

BLAST

4 Extract consensus HE sequence from the MAFFT aligned 192 whole genomes (see below) using Geneious

Discontiguous MegaBLAST input: **HE_realigned_consensus_sequence.fasta**



Database: Nucleotide collection (nr/nt) (AA or

Add/Remove Databases

Program: Discontiguous Megablast - slower,

Results: Hit table

Retrieve: Matching region with annotations (sl...

Maximum Hits: 1,000

☑ Low Complexity Filter

☑ Mask for lookup table

□ Human Repeats Filter

Max E-value: 1e-20

Word Size: 11

Gap cost (Open Extend): 5 2

Scoring (Match Mismatch): 2 -3

Max Target Seqs: 500

Template length: 16

Template Type: Maximal

Entrez Query: BetaCoronaVirus[Organism]

Other Arguments:

Discontiguous MegaBLAST of BCoV HE consensus sequence with E-value of E-20 against all BetaCoronavirus sequences in the GenBank nt database

Output: **BetaCoronaVirus_HE_BlastMatch_Bovine_HE_Consensus.txt** This list of 844 subject sequences (hits) was reduced to 714 by requiring that each hit covered at least 98.4% (1255 bp) of query HE sequence resulting in the following FASTA file :

BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus.fasta

This file was filtered for sequences with "bovine coronavirus" in the FASTA description field yielding :

345_Sequences_Bovine_HE_only.fasta

MAFFT Alignment

5 Align using MAFFT v7.450

The sequences were aligned using the MAFFT accuracy methods ---globalpair as defined in the mafft manual page below

DESCRIPTION
MAFFT is a multiple sequence alignment program for unix-like operating systems. It offers a range of multiple alignment methods.

Accuracy-oriented methods:
• L-INS-i (probably most accurate; recommended for <200 sequences; iterative refinement method incorporating local pairwise alignment information):

mafft --localpair --maxiterate 1000input [> output]

linsi input [> output]

• G-INS-i (suitable for sequences of similar lengths; recommended for <200 sequences; iterative refinement method incorporating global pairwise alignment
information):

mafft--globalpair--maxiterate1000input [> output]

ginsi input [> output]

Keep in mind ...

1. The commands below generate files that contain both STDOUT (mafft output to the "terminal") as well as the multifasta alignment file (.afa). The STDOUT is found at the beginning of the file and the alignments follow the STDOUT
2. All .afa_out files were split in a text editor into two separate .out and .afa files for downstream analysis.
3. The input FASTA file headers contain description information; this header information should be stripped out to facilitate readable, compact leaf names/identifiers. Use the following command to strip out the description text.

sed to strip out the description in the header from FASTA

```
sed '/^>/ s/ .*//'  input_alignment_multifasta_file.afa  >
output_alignment_multifasta_file.SEQID_Only.afa
```

Strip out description text from multifasta alignment file
UBUNTU Linux

5.1 **The input FASTA files are:**

192_Spike_6_14_22.fasta **192_HE_6_14_22.fasta**
192_Genomes_Final_6_14_22.fasta **345_Sequences_Bovine_HE_only.fasta**
BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus.fasta
HE_realigned_consensus_sequence.fasta

The spike and HE genes were extracted from the annotated genomes into the file of fasta sequences above. There are duplicated sequences within 192_HE_6_22.fasta as well as within 192_Spike_6_14_22.fasta. The genome sequences within 192_Genomes_Final_6_14_22.fasta are unique.

The spike and HE FASTA sequence files were deduplicated with **fassort** & **fasuniq** from the FAST Analysis of Sequences Toolbox. The unique sequences present in multiple genomes will be designated with a sequence identifier comprised of a concatenation of the sequence identifiers used in the multiple genomes separated by a "__" or ":". For example, the HE fasta sequence id IWT-11:SHG-6:TCG-21:TCG-23:TCG-22 represents the identical HE fasta sequences IWT-11, SHG-6, TCG-21, TCG-23 and TCG-22.

Create file of unique spike FASTA sequences

```
fassort -s 192_HE_6_14_22.fasta | fasuniq --concat='__'  >
unique_HE_6_14_22.fasta
```

Create a file of unique HE FASTA sequences. The unique sequences present in multiple genomes will be designated with a sequence identifier comprised of a concatenation of the sequence identifiers used in the multiple genomes, separated by a double underscore "__"
UBUNTU Linux

unique_HE_6_14_22.fasta

Citation: Aspen M Workman, Tara G. McDaneld, Gregory P Harhay, Subha Das, John Dustin Loy, Benjamin M. Hause Phylogenetic Analysis of Complete Bovine Coronavirus Genome Sequences https://dx.doi.org/10.17504/protocols.io.kqdg3pyeql25/v3

Create file of unique Spike FASTA sequences

```
fassort -s 192_Spike_6_14_22.fasta | fasuniq --concat='__'  >
unique_Spike_6_14_22.fasta
```

Create a file of unique Spike FASTA sequences. The unique sequences present in multiple genomes will be designated with a sequence identifier comprised of a concatenation of the sequence identifiers used in the multiple genomes, separated by a double underscore "__"
UBUNTU Linux

unique_Spike_6_14_22.fasta

```
fassort -s 345_Sequences_Bovine_HE_only.fasta | fasuniq --concat='__' >
unique_345_Sequences_Bovine_HE_only.fasta
```

unique_345_Sequences_Bovine_HE_only.fasta

```
fassort -s
BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus.fasta |
fasuniq --concat="__" >
unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus
```

unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus.fasta

5.2 Remove description text (all text after sequence id) in HE , Spike, and Genomes fasta fileps.

sed to strip out the description in the header from FASTA

```
sed '/^>/ s/ .*//'  unique_HE_6_14_22.fasta  >
unique_HE_6_14_22.SEQID_Only.fasta
```

UBUNTU Linux

output : unique_HE_6_14_22.SEQID_Only.fasta

sed to strip out the description in the header from FASTA

```
sed '/^>/ s/ .*//'  unique_HE_6_14_22.fasta  >
unique_HE_6_14_22.SEQID_Only.fasta
```

UBUNTU Linux

output: unique_Spike_6_14_22.SEQID_Only.fasta

sed to strip out the description in the header from FASTA

```
sed '/^>/ s/ .*//'  192_Genomes_Final_6_14_22.fasta >
192_Genomes_Final_6_14_22.SEQID_Only.fasta
```

UBUNTU Linux

output: 192_Genomes_Final_6_14_22.SEQID_Only.fasta

```
sed '/^>/ s/ .*//' unique_345_Sequences_Bovine_HE_only.fasta  >
unique_345_Sequences_Bovine_HE_only.SEQID_Only.fasta
```

unique_345_Sequences_Bovine_HE_only.SEQID_Only.fasta

```
sed '/^>/ s/ .*//'
unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus
unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus
```

unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus_SeqID_Only.fasta

5.3 Align 192 **BCV** genomes using the default **PAM (JTT)** 200 substitution matrix

```
nohup mafft --thread 25 --globalpair --maxiterate 1000 --jtt 200 --reorder
192_Genomes_Final_6_14_22.SEQID_Only.fasta >
192_Genomes_Final_6_14_22.SEQID_Only.jtt_200.globalpair.afa_out &
```

outputs : 192_Genomes_Final_6_14_22.jtt_200.globalpair.SEQID_Only.afa_out

This was split into 192_Genomes_Final_6_14_22.jtt_200.globalpair.SEQID_Only.afa &
192_Genomes_Final_6_14_22.jtt_200.globalpair.SEQID_Only.out

5.4 Align **192 BCV genomes** using the **PAM (JTT) 100** substitution matrix

```
nohup mafft --thread 30 --globalpair --maxiterate 1000 --jtt 100 --reorder
192_Genomes_Final_6_14_22.SEQID_Only.fasta >
192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.afa_out &
```

outputs : **192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.afa_out**

This was split into **192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.afa** & **192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.out**

5.5 Align **SPIKE** using the default **PAM (JTT) 200** substitution matrix

```
nohup mafft --thread 15 --globalpair --maxiterate 1000 --jtt 200 --reorder
unique_Spike_6_14_22.SEQID_Only.fasta >
unique_Spike_6_14_22.SEQID_Only.jtt_200.globalpair.afa_out &
```

Outputs : **unique_Spike_6_14_22.SEQID_Only.jtt_200.globalpair.afa_out**

This was split into **unique_Spike_6_14_22.SEQID_Only.jtt_200.globalpair.afa** & **unique_Spike_6_14_22.SEQID_Only.jtt_200.globalpair.out**

5.6 Align **SPIKE** using the **PAM (JTT) 100** substitution matrix

```
nohup mafft --thread 15 --globalpair --maxiterate 1000 --jtt 100 --reorder
unique_Spike_6_14_22.SEQID_Only.fasta >
unique_Spike_6_14_22.SEQID_Only.jtt_100.globalpair.afa_out &
```

Outputs : **unique_Spike_6_14_22.SEQID_Only.jtt_100.globalpair.afa_out**

This was split into **unique_Spike_6_14_22.SEQID_Only.jtt_100.globalpair.afa** & **unique_Spike_6_14_22.SEQID_Only.jtt_100.globalpair.out**

5.7 Align **HE** using the default **PAM (JTT) 200** substitution matrix

```
nohup mafft --thread 15 --globalpair --maxiterate 1000 --jtt 200 --reorder
unique_HE_6_14_22.SEQID_Only.fasta >
unique_HE_6_14_22.SEQID_Only.jtt_200.globalpair.afa_out
```

Outputs : **unique_HE_6_14_22.SEQID_Only.jtt_200.globalpair.afa_out**

This was split into **unique_HE_6_14_22.SEQID_Only.jtt_200.globalpair.afa** & **unique_HE_6_14_22.SEQID_Only.jtt_200.globalpair.out**

5.8 Align **HE** using the **PAM (JTT) 100** substitution matrix

```
nohup mafft --thread 15 --globalpair --maxiterate 1000 --jtt 100 --reorder
unique_HE_6_14_22.SEQID_Only.fasta >
unique_HE_6_14_22.SEQID_Only.jtt_100.globalpair.afa_out
```

Output: **unique_HE_6_14_22.SEQID_Only.jtt_100.globalpair.afa_out**

This was split into: **unique_HE_6_14_22.SEQID_Only.jtt_100.globalpair.afa** & **unique_HE_6_14_22.SEQID_Only.jtt_100.globalpair.out**

5.9 Align unique BCV HE sequences the **PAM (JTT) 100** substitution matrix

```
nohup mafft --thread 25 --globalpair --maxiterate 1000 --jtt 100 --reorder
unique_345_Sequences_Bovine_HE_only.SEQID_Only.fasta>
unique_345_Sequences_Bovine_HE_only.SEQID_Only.jtt_100.globalpair.afa_o
&
```

Output:
unique_345_Sequences_Bovine_HE_only.SEQID_Only.jtt_100.globalpair.afa_out

Was split into
unique_345_Sequences_Bovine_HE_only.SEQID_Only.jtt_100.globalpair.afa &
unique_345_Sequences_Bovine_HE_only.SEQID_Only.jtt_100.globalpair.out

```
nohup mafft --thread 15 --globalpair --maxiterate 1000 --jtt 100 --reorder
unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus
unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus
```

unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus_SeqID_Only.jtt_100.globalpair.afa_out

unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus_SeqID_Only.jtt_100.globalpair.afa

unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus_SeqID_Only.jtt_100.globalpair.out

Tree Building

6 RaXML Tee Building

Notes below are cut from the RaXML manual to put the command line parameters used in context

> "This is a how-to, which describes how RAxML should best be used for a simple real-world biological analysis, given an example alignment named ex_al"
>
> Now, if you want to run a full analysis, i.e., BS and ML search type:
>
> ```
> raxmlHPC -f a -x 12345 -p 12345 -# 100 m GTRGAMMA -s ex_al -n TEST
> ```
>
> This will first conduct a BS search and once that is done a search for the best–scoring ML tree.
> Such a program run will return the bootstrapped trees (RAxML_bootstrap.TEST), the best scoring ML tree(RAxML_bestTree.TEST), and the BS support values drawn on the best-scoring tree as node labels (RAxML_bipartitions.TEST) as well as, more correctly since support values refer to branches as branch labels (RAxML_bipartitionsBranchLabels.TEST).
>
> Finally, note that by increasing the number of BS replicates via -# you will also make the ML thorough since for ML optimization every 5th BS tree is used as a starting point to
> for ML trees.

raxmlHPC reports 192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.afa has 3438 DNA alignment patterns

> Thus, if you run RAxML with 32 instead of 1 thread this does not mean that it will automatically become 32 times faster, it may actually even become slower. As I already mentioned, the parallel efficiency, that is, with how many threads/cores you can still execute it efficiently in parallel depends on the alignment length, or to be more precise on the number of distinct patterns in your alignment.
>
> This number is printed by RAxML to the terminal and into the RAxML_info.runID file"
> and look like this:
>
> Alignment has 70 distinct alignment patterns
>
> As a rule of thumb **I'd use one core/thread per 500 DNA site patterns**, i.e., if you have less, than it's probably better to just use the sequential version. Single-gene DNA alignments with around 1000 sites can be analyzed with 2 or at most 4 threads. Thus, the more patterns your alignment has, the more threads/cores you can use efficiently.

Given the directions above from the RaXML manual, use 8 threads for raxmlHPC tree building

```
nohup raxmlHPC -T 8 -f a -x 12345 -p 12345 -N 2000 -m GTRGAMMA -s
192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.afa -n
192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000 >
192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000.out &
```

Run raxmlHPC command using nohup on 'nix platform to protect from premature job termination in case of remote connection loss ... and run in background with & at end of command line. Perform a 2000 boostrap search followed by a search for the best-scoring maximum likelihood tree.

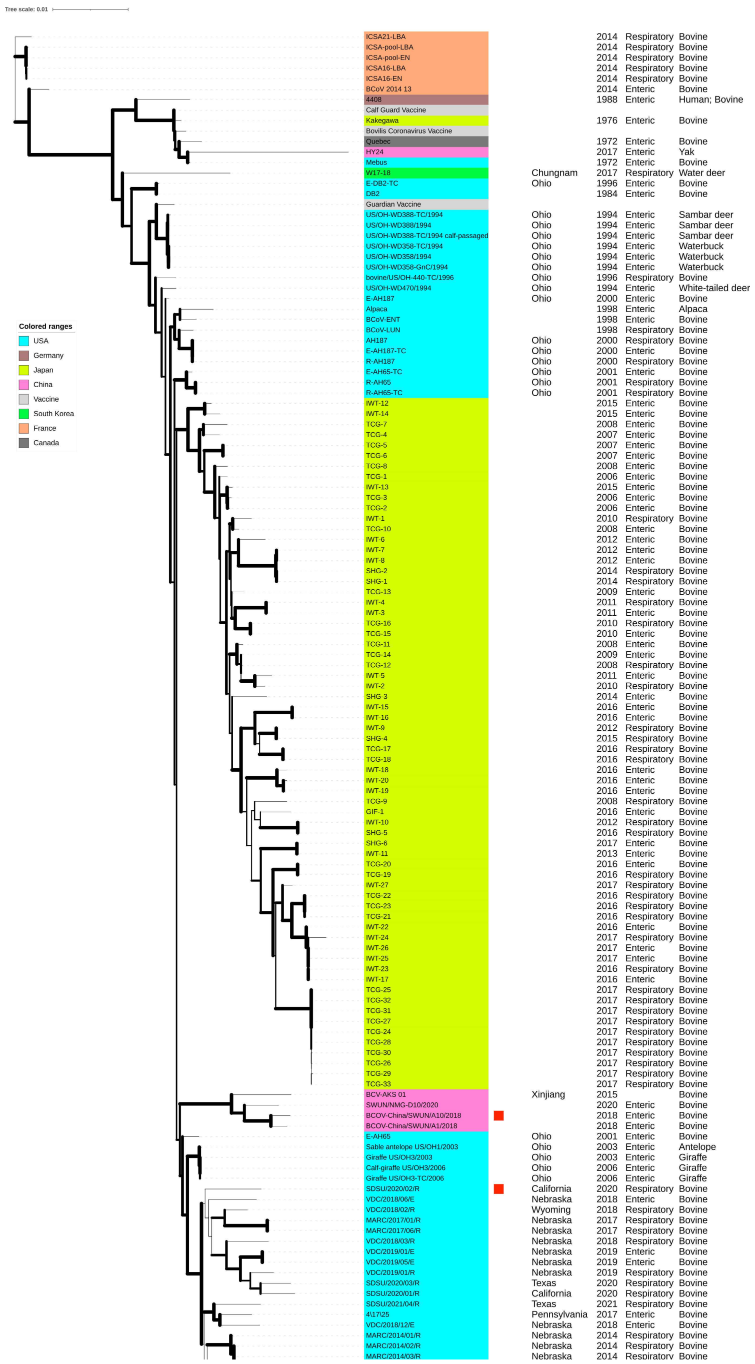
-T 8 (use 8 threads)
-f a
-x 12345
-p 12345
-m GTRGAMMA
-N 2000

Output :

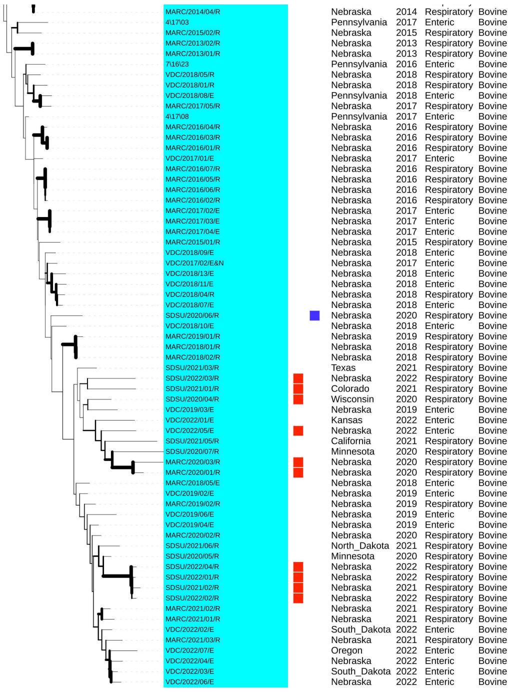
- RAxML_bootstrap.192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000
- 192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000.out
- RAxML_info.192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000
- RAxML_bipartitionsBranchLabels.192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000
- RAxML_bipartitions.192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000
- RAxML_bestTree.192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000

Use RAxML_bipartitionsBranchLabels.192_Genomes_Final_6_14_22.jtt_100.globalpair.SEQID_Only.T8.N2000 file to create phylogenetic tree with bootstrap support values provided as branch labels.

Visualize this tree of 192 Bovine coronavirus genomes, MAFFT Global Pair Alignment, PAM 100 in the Interactive Tree of Life


Tree scale: 0.01
Colored ranges
USA
Germany
Japan
China
Vaccine
South Korea
France
Canada
ICSA21-LBA 2014 Respiratory Bovine
ICSA-pool-LBA 2014 Respiratory Bovine
ICSA-pool-EN 2014 Respiratory Bovine
ICSA16-LBA 2014 Respiratory Bovine
ICSA16-EN 2014 Respiratory Bovine
BCoV 2014 13 2014 Enteric Bovine
4408 1988 Enteric Human; Bovine
Calf Guard Vaccine
Kakegawa 1976 Enteric Bovine
Bovilis Coronavirus Vaccine
Quebec 1972 Enteric Bovine
HY24 2017 Enteric Yak
Mebus 1972 Enteric Bovine
W17-18 Chungnam 2017 Respiratory Water deer
E-DB2-TC Ohio 1996 Enteric Bovine
DB2 1984 Enteric Bovine
Guardian Vaccine
US/OH-WD388-TC/1994 Ohio 1994 Enteric Sambar deer
US/OH-WD388/1994 Ohio 1994 Enteric Sambar deer
US/OH-WD388-TC/1994 calf-passaged Ohio 1994 Enteric Sambar deer
US/OH-WD358-TC/1994 Ohio 1994 Enteric Waterbuck
US/OH-WD358/1994 Ohio 1994 Enteric Waterbuck
US/OH-WD358-GnC/1994 Ohio 1994 Enteric Waterbuck
bovine/US/OH-440-TC/1996 Ohio 1996 Respiratory Bovine
US/OH-WD470/1994 Ohio 1994 Enteric White-tailed deer
E-AH187 Ohio 2000 Enteric Bovine
Alpaca 1998 Enteric Alpaca
BCoV-ENT 1998 Enteric Bovine
BCoV-LUN 1998 Respiratory Bovine
AH187 Ohio 2000 Respiratory Bovine
E-AH187-TC Ohio 2000 Enteric Bovine
R-AH187 Ohio 2000 Respiratory Bovine
E-AH65-TC Ohio 2001 Enteric Bovine
R-AH65 Ohio 2001 Respiratory Bovine
R-AH65-TC Ohio 2001 Respiratory Bovine
IWT-12 2015 Enteric Bovine
IWT-14 2015 Enteric Bovine
TCG-7 2008 Enteric Bovine
TCG-4 2007 Enteric Bovine
TCG-5 2007 Enteric Bovine
TCG-6 2007 Enteric Bovine
TCG-8 2008 Enteric Bovine
TCG-1 2006 Enteric Bovine
IWT-13 2015 Enteric Bovine
TCG-3 2006 Enteric Bovine
TCG-2 2006 Enteric Bovine
IWT-1 2010 Respiratory Bovine
TCG-10 2008 Enteric Bovine
IWT-6 2012 Enteric Bovine
IWT-7 2012 Enteric Bovine
IWT-8 2012 Enteric Bovine
SHG-2 2014 Respiratory Bovine
SHG-1 2014 Respiratory Bovine
TCG-13 2009 Enteric Bovine
IWT-4 2011 Respiratory Bovine
IWT-3 2011 Enteric Bovine
TCG-16 2010 Respiratory Bovine
TCG-15 2010 Enteric Bovine
TCG-11 2008 Enteric Bovine
TCG-14 2009 Enteric Bovine
TCG-12 2008 Respiratory Bovine
IWT-5 2011 Enteric Bovine
IWT-2 2010 Respiratory Bovine
SHG-3 2014 Enteric Bovine
IWT-15 2016 Enteric Bovine
IWT-16 2016 Enteric Bovine
IWT-9 2012 Respiratory Bovine
SHG-4 2015 Respiratory Bovine
TCG-17 2016 Respiratory Bovine
TCG-18 2016 Respiratory Bovine
IWT-18 2016 Enteric Bovine
IWT-20 2016 Enteric Bovine
IWT-19 2016 Enteric Bovine
TCG-9 2008 Respiratory Bovine
GIF-1 2016 Enteric Bovine
IWT-10 2012 Respiratory Bovine
SHG-5 2016 Respiratory Bovine
SHG-6 2017 Enteric Bovine
IWT-11 2013 Enteric Bovine
TCG-20 2016 Enteric Bovine
TCG-19 2016 Respiratory Bovine
IWT-27 2017 Respiratory Bovine
TCG-22 2016 Respiratory Bovine
TCG-23 2016 Respiratory Bovine
TCG-21 2016 Respiratory Bovine
IWT-22 2016 Enteric Bovine
IWT-24 2017 Respiratory Bovine
IWT-26 2017 Enteric Bovine
IWT-25 2017 Enteric Bovine
IWT-23 2016 Respiratory Bovine
IWT-17 2016 Enteric Bovine
TCG-25 2017 Respiratory Bovine
TCG-32 2017 Respiratory Bovine
TCG-31 2017 Respiratory Bovine
TCG-27 2017 Respiratory Bovine
TCG-24 2017 Respiratory Bovine
TCG-28 2017 Respiratory Bovine
TCG-30 2017 Respiratory Bovine
TCG-26 2017 Respiratory Bovine
TCG-29 2017 Respiratory Bovine
TCG-33 2017 Respiratory Bovine
BCV-AKS 01 Xinjiang 2015 Bovine
SWUN/NMG-D10/2020 2020 Enteric Bovine
BCOV-China/SWUN/A10/2018 🟥 2018 Enteric Bovine
BCOV-China/SWUN/A1/2018 2018 Enteric Bovine
E-AH65 Ohio 2001 Enteric Bovine
Sable antelope US/OH1/2003 Ohio 2003 Enteric Antelope
Giraffe US/OH3/2003 Ohio 2003 Enteric Giraffe
Calf-giraffe US/OH3/2006 Ohio 2006 Enteric Giraffe
Giraffe US/OH3-TC/2006 Ohio 2006 Enteric Giraffe
SDSU/2020/02/R 🟥 California 2020 Respiratory Bovine
VDC/2018/06/E Nebraska 2018 Enteric Bovine
VDC/2018/02/R Wyoming 2018 Respiratory Bovine
MARC/2017/01/R Nebraska 2017 Respiratory Bovine
MARC/2017/06/R Nebraska 2017 Respiratory Bovine
VDC/2018/03/R Nebraska 2018 Respiratory Bovine
VDC/2019/01/E Nebraska 2019 Enteric Bovine
VDC/2019/05/E Nebraska 2019 Enteric Bovine
VDC/2019/01/R Nebraska 2019 Respiratory Bovine
SDSU/2020/03/R Texas 2020 Respiratory Bovine
SDSU/2020/01/R California 2020 Respiratory Bovine
SDSU/2021/04/R Texas 2021 Respiratory Bovine
4\17\25 Pennsylvania 2017 Enteric Bovine
VDC/2018/12/E Nebraska 2018 Enteric Bovine
MARC/2014/01/R Nebraska 2014 Respiratory Bovine
MARC/2014/02/R Nebraska 2014 Respiratory Bovine
MARC/2014/03/R Nebraska 2014 Respiratory Bovine

Citation: Aspen M Workman, Tara G. McDaneld, Gregory P Harhay, Subha Das, John Dustin Loy, Benjamin M. Hause Phylogenetic Analysis of Complete Bovine Coronavirus Genome Sequences https://dx.doi.org/10.17504/protocols.io.kqdg3pyeql25/v3


MARC/2014/04/R Nebraska 2014 Respiratory Bovine
4\17\03 Pennsylvania 2017 Enteric Bovine
MARC/2015/02/R Nebraska 2015 Respiratory Bovine
MARC/2013/02/R Nebraska 2013 Respiratory Bovine
MARC/2013/01/R Nebraska 2013 Respiratory Bovine
7\16\23 Pennsylvania 2016 Enteric Bovine
VDC/2018/05/R Nebraska 2018 Respiratory Bovine
VDC/2018/01/R Nebraska 2018 Respiratory Bovine
VDC/2018/08/E Pennsylvania 2018 Enteric Bovine
MARC/2017/05/R Nebraska 2017 Respiratory Bovine
4\17\08 Pennsylvania 2017 Enteric Bovine
MARC/2016/04/R Nebraska 2016 Respiratory Bovine
MARC/2016/03/R Nebraska 2016 Respiratory Bovine
MARC/2016/01/R Nebraska 2016 Respiratory Bovine
VDC/2017/01/E Nebraska 2017 Enteric Bovine
MARC/2016/07/R Nebraska 2016 Respiratory Bovine
MARC/2016/05/R Nebraska 2016 Respiratory Bovine
MARC/2016/06/R Nebraska 2016 Respiratory Bovine
MARC/2016/02/R Nebraska 2016 Respiratory Bovine
MARC/2017/02/E Nebraska 2017 Enteric Bovine
MARC/2017/03/E Nebraska 2017 Enteric Bovine
MARC/2017/04/E Nebraska 2017 Enteric Bovine
MARC/2015/01/R Nebraska 2015 Respiratory Bovine
VDC/2018/09/E Nebraska 2018 Enteric Bovine
VDC/2017/02/E&N Nebraska 2017 Enteric Bovine
VDC/2018/13/E Nebraska 2018 Enteric Bovine
VDC/2018/11/E Nebraska 2018 Enteric Bovine
VDC/2018/04/R Nebraska 2018 Respiratory Bovine
VDC/2018/07/E Nebraska 2018 Enteric Bovine
SDSU/2020/06/R 🟦 Nebraska 2020 Respiratory Bovine
VDC/2018/10/E Nebraska 2018 Enteric Bovine
MARC/2019/01/R Nebraska 2019 Respiratory Bovine
MARC/2018/01/R Nebraska 2018 Respiratory Bovine
MARC/2018/02/R Nebraska 2018 Respiratory Bovine
SDSU/2021/03/R Texas 2021 Respiratory Bovine
SDSU/2022/03/R 🟥 Nebraska 2022 Respiratory Bovine
SDSU/2021/01/R 🟥 Colorado 2021 Respiratory Bovine
SDSU/2020/04/R 🟥 Wisconsin 2020 Respiratory Bovine
VDC/2019/03/E Nebraska 2019 Enteric Bovine
VDC/2022/01/E Kansas 2022 Enteric Bovine
VDC/2022/05/E 🟥 Nebraska 2022 Enteric Bovine
SDSU/2021/05/R California 2021 Respiratory Bovine
SDSU/2020/07/R Minnesota 2020 Respiratory Bovine
MARC/2020/03/R 🟥 Nebraska 2020 Respiratory Bovine
MARC/2020/01/R 🟥 Nebraska 2020 Respiratory Bovine
MARC/2018/05/E Nebraska 2018 Enteric Bovine
VDC/2019/02/E Nebraska 2019 Enteric Bovine
MARC/2019/02/R Nebraska 2019 Respiratory Bovine
VDC/2019/06/E Nebraska 2019 Enteric Bovine
VDC/2019/04/E Nebraska 2019 Enteric Bovine
MARC/2020/02/R Nebraska 2020 Respiratory Bovine
SDSU/2021/06/R North_Dakota 2021 Respiratory Bovine
SDSU/2020/05/R Minnesota 2020 Respiratory Bovine
SDSU/2022/04/R 🟥 Nebraska 2022 Respiratory Bovine
SDSU/2022/01/R 🟥 Nebraska 2022 Respiratory Bovine
SDSU/2021/02/R 🟥 Nebraska 2021 Respiratory Bovine
SDSU/2022/02/R 🟥 Nebraska 2022 Respiratory Bovine
MARC/2021/02/R Nebraska 2021 Respiratory Bovine
MARC/2021/01/R Nebraska 2021 Respiratory Bovine
VDC/2022/02/E South_Dakota 2022 Enteric Bovine
MARC/2021/03/R Nebraska 2021 Respiratory Bovine
VDC/2022/07/E Oregon 2022 Enteric Bovine
VDC/2022/04/E Nebraska 2022 Enteric Bovine
VDC/2022/03/E South_Dakota 2022 Enteric Bovine
VDC/2022/06/E Nebraska 2022 Enteric Bovine

RAxML tree of 192 bovine coronavirus genomes. Bootstrap support values are proportional to the line thickness of the branches. Isolates with HE deletions are denoted with a filled red square while the single isolate with an HE insertion is denoted with a filled blue square. Tree metrics are best investigated in the IToL tree.

The pdf suitable for download is RaXML tree of 192 livestock coronavirus genomes.pdf

7 Following the same approach with the 192 BCoV genomes other RaXML Trees were generated for spike and HE and are included here

□ RAxML_bipartitionsBranchLabels.unique_HE_6_14_22.SEQID_Only.jtt_100.globalpair.T8.N2000

□ RAxML_bipartitionsBranchLabels.unique_HE_6_14_22.SEQID_Only.jtt_200.globalpair.T8.N2000

□ RAxML_bipartitionsBranchLabels.unique_Spike_6_14_22.SEQID_Only.jtt_100.globalpair.T8.N2000

□ RAxML_bipartitionsBranchLabels.unique_Spike_6_14_22.SEQID_Only.jtt_200.globalpair.T8.N2000

□ RAxML_bipartitionsBranchLabels.unique_345_Sequences_Bovine_HE_only.SEQID_Only.jtt_100.globalpair.T8.N2000

□ unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus_SeqID_Only.T8.N200.out

□ RAxML_bipartitionsBranchLabels.unique_BetaCoronaVirus_HE_GE_1255_BP_BlastMatch_Bovine_HE_Consensus_SeqID_Only.T8.N2000

Citation: Aspen M Workman, Tara G. McDaneld, Gregory P Harhay, Subha Das, John Dustin Loy, Benjamin M. Hause Phylogenetic Analysis of Complete Bovine Coronavirus Genome Sequences https://dx.doi.org/10.17504/protocols.io.kqdg3pyeql25/v3