

# concatenated\_host\_viral

FRIDAY, 4/5/2019

## Concatenating host and viral genome

```
cat Macaca_fascicularis.Macaca_fascicularis_5.0.dna.toplevel.fa Makona_genome.fasta > Cynomolgus_Makona.fasta
```

## Open fasta file

```
nano Cynomolgus_Makona.fasta
```

- Scroll to end and find where Makona genome starts

### Screen Shot 2018-12-04 at 10.30.56 AM.png

```

-AQIA01082004.1 dna:scaffold scaffold:Macaca_fascicularis_5.0:AQIA01082004.1:1:248:1 REF
CTCGGAACCTCTCTCAGATATCTATCTGAAGATTTCCATAAGATATTTCCATTTC
ACCATAGCTTCAAAGGCTCCGAATATCTTTCCGAGACTAAACATAAACAGTGTAG
AAAACCTGCTCAATGAAAGAGAGCTTACCTGCAGCAGAGAAATTCACACAGCTCAATGA
AAAAGAAAGCTTAACCTGAGCAGAGGAATTCACATGATCAGTAGAGTTTCACAGAAAGC
TTCTCTCT
-AQIA01081996.1 dna:scaffold scaffold:Macaca_fascicularis_5.0:AQIA01081996.1:1:218:1 REF
TAATAAAGAGGGAATTTATAGAAGATACAGAGTTTTCGAGAGTTCCACAGCT
GCGAGGGTCAACAGCGAGAATCACGCTTTCTGAAGCAGAGATATCTCTAACACAA
ACTCTTAATCTCGCTTCGACCATCAGGAGAGAAATCTTAGAGGACAGTTCTCCTT
GTATACGATAAGGACTAGAATGTGGATTCTTTCT
-gb:K160347.1 Organism:Zaire ebolavirus Strain Name:H.sapiens-wt/GIN/2014 Makona-Gueckedou-C07 Segment:null Host:Human
CGGACACAAAAAGAGGAATTTTAGGATCTTTGTGTGGAATACTAGGAGATTATATA
TTTTCTCTCATTGAAATTTATATCGGAATTTAAATTTGAATTTGTTACTGTAATCATACCTGGTTGTTT
CAGAGCATATCACCAGATAGAGAACCTAGGCTCCGAGAGGGCAGGGCATCAGTGTGCTCAGT
TGAATTCCTTGTGCAATCTAGGCTTATCAATCAGAGTCCGCTTAACTCTCGAGGGTGATCC

```

- annotate header of the Makona genome to look similar to Cynomolgus
- choose an identifier that follows Cynomolgus format but is also unique from all other identifiers (use command/ctrl find)

### Screen Shot 2018-12-04 at 10.32.51 AM.png

```

-AQIA01082004.1 dna:scaffold scaffold:Macaca_fascicularis_5.0:AQIA01082004.1:1:248:1 REF
CTCGGAACCTCTCTCAGATATCTATCTGAAGATTTCCATAAGATATTTCCATTTC
ACCATAGCTTCAAAGGCTCCGAATATCTTTCCGAGACTAAACATAAACAGTGTAG
AAAACCTGCTCAATGAAAGAGAGCTTACCTGCAGCAGAGAAATTCACACAGCTCAATGA
AAAAGAAAGCTTAACCTGAGCAGAGGAATTCACATGATCAGTAGAGTTTCACAGAAAGC
TTCTCTCT
-AQIA01081996.1 dna:scaffold scaffold:Macaca_fascicularis_5.0:AQIA01081996.1:1:218:1 REF
TAATAAAGAGGGAATTTATAGAAGATACAGAGTTTTCGAGAGTTCCACAGCT
GCGAGGGTCAACAGCGAGAATCACGCTTTCTGAAGCAGAGATATCTCTAACACAA
ACTCTTAATCTCGCTTCGACCATCAGGAGAGAAATCTTAGAGGACAGTTCTCCTT
GTATACGATAAGGACTAGAATGTGGATTCTTTCT
-gb:K160347.1 Organism:Zaire ebolavirus Strain Name:H.sapiens-wt/GIN/2014 Makona-Gueckedou-C07 Segment:null Host:Human
CGGACACAAAAAGAGGAATTTTAGGATCTTTGTGTGGAATACTAGGAGATTATATA
TTTTCTCTCATTGAAATTTATATCGGAATTTAAATTTGAATTTGTTACTGTAATCATACCTGGTTGTTT

```

-Final should look like this

### Screen Shot 2018-12-04 at 10.33.13 AM.png

```

-AQIA01082004.1 dna:scaffold scaffold:Macaca_fascicularis_5.0:AQIA01082004.1:1:248:1 REF
CTCGGAACCTCTCTCAGATATCTATCTGAAGATTTCCATAAGATATTTCCATTTC
ACCATAGCTTCAAAGGCTCCGAATATCTTTCCGAGACTAAACATAAACAGTGTAG
AAAACCTGCTCAATGAAAGAGAGCTTACCTGCAGCAGAGAAATTCACACAGCTCAATGA
AAAAGAAAGCTTAACCTGAGCAGAGGAATTCACATGATCAGTAGAGTTTCACAGAAAGC
TTCTCTCT
-AQIA01081996.1 dna:scaffold scaffold:Macaca_fascicularis_5.0:AQIA01081996.1:1:218:1 REF
TAATAAAGAGGGAATTTATAGAAGATACAGAGTTTTCGAGAGTTCCACAGCT
GCGAGGGTCAACAGCGAGAATCACGCTTTCTGAAGCAGAGATATCTCTAACACAA
ACTCTTAATCTCGCTTCGACCATCAGGAGAGAAATCTTAGAGGACAGTTCTCCTT
GTATACGATAAGGACTAGAATGTGGATTCTTTCT
-AQIA99999999.1 dna:chromosome chromosome:Macaca_fascicularis_5.0:AQIA99999999.1:1:18959:1 REF
CGGACACAAAAAGAGGAATTTTAGGATCTTTGTGTGGAATACTAGGAGATTATATA
TTTTCTCTCATTGAAATTTATATCGGAATTTAAATTTGAATTTGTTACTGTAATCATACCTGGTTGTTT

```

## Open GTF file and add additional information at the end.

- Manually annotate information about the viral genome/gene
- Make sure to follow Cynomolgus macaque format (9 columns of information)
- based on whether you are annotating the viral gene as a gene, exon or CDS, follow the same format of the last column

Screen Shot 2018-12-04 at 10.36.25 AM.png

Accession	Organism	Gene	Transcript	Exon	Start	End	Strand	Feature
U010184.1	ensembl	gene	361	462	-	-	-	gene_id "ENSGMAG00000000052"; gene_version "1"; gene_name "EIF00019"; gene_source "ensembl";
U010184.1	ensembl	transcript	361	462	-	-	-	gene_id "ENSGMAG00000000052"; gene_version "1"; transcript_id "ENSTGAT00000000427";
U010184.1	ensembl	exon	361	462	-	-	-	gene_id "ENSGMAG00000000052"; gene_version "1"; transcript_id "ENSTGAT00000000427"; transcript
U019999999.1	ensembl	gene	1	18959	-	-	-	gene_id "EB00"; gene_version "1"; gene_name "EB00.gnome"; gene_source "ensembl";
U019999999.1	ensembl	transcript	1	18959	-	-	-	gene_id "EB00"; gene_version "1"; transcript_id "EB00.1"; transcript_source "ensembl";
U019999999.1	ensembl	exon	1	18959	-	-	-	gene_id "EB00"; gene_version "1"; transcript_id "EB00.1"; exon_number "1"; exon

- Final should look like this
- Save as a new name that incorporates viral name.

Screen Shot 2018-12-04 at 10.36.52 AM.png

hg1808146.1	ensembl	gene	361	462	gene_id "ENSGA0000000079"; gene_version "1"; gene_name "IFI0019"; gene_source "ensembl"; gene_biotype "l
hg1808146.1	ensembl	transcript	361	462	gene_id "ENSGA0000000079"; gene_version "1"; transcript_id "ENSGA000000040127"; transcript_ver
hg1808146.1	ensembl	exon	361	462	gene_id "ENSGA0000000079"; gene_version "1"; transcript_id "ENSGA000000040127"; transcript_ver
hg1899999.1	ensembl	gene	1	18999	gene_id "EBV"; gene_version "1"; gene_name "EBV_gene"; gene_source "ensembl"; gene_biotype "protein_cod
hg1899999.1	ensembl	exon	1	18999	gene_id "EBV"; gene_version "1"; transcript_id "EBV"; gene_source "ensembl"; exon_number "1"; gene_nu