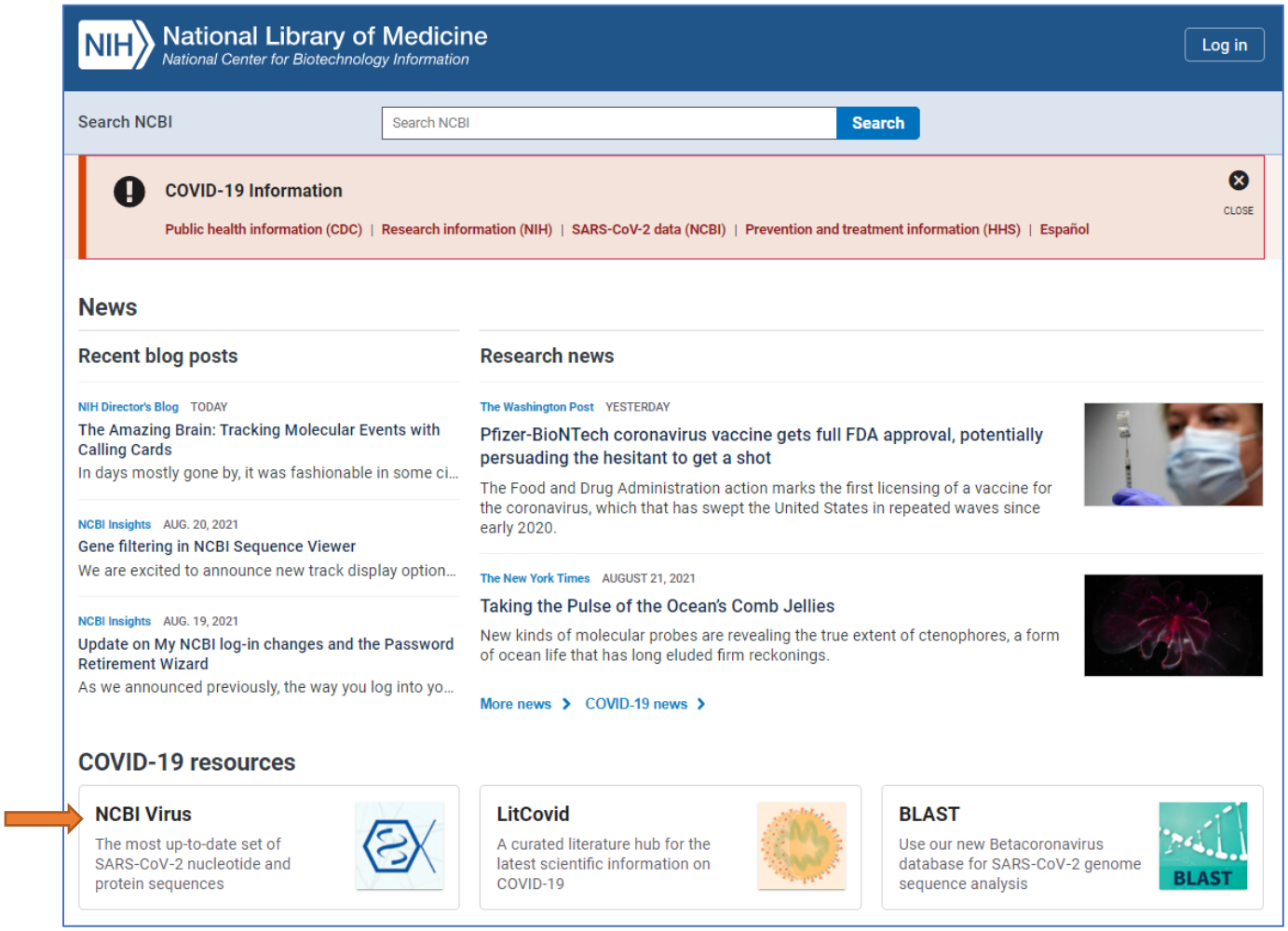
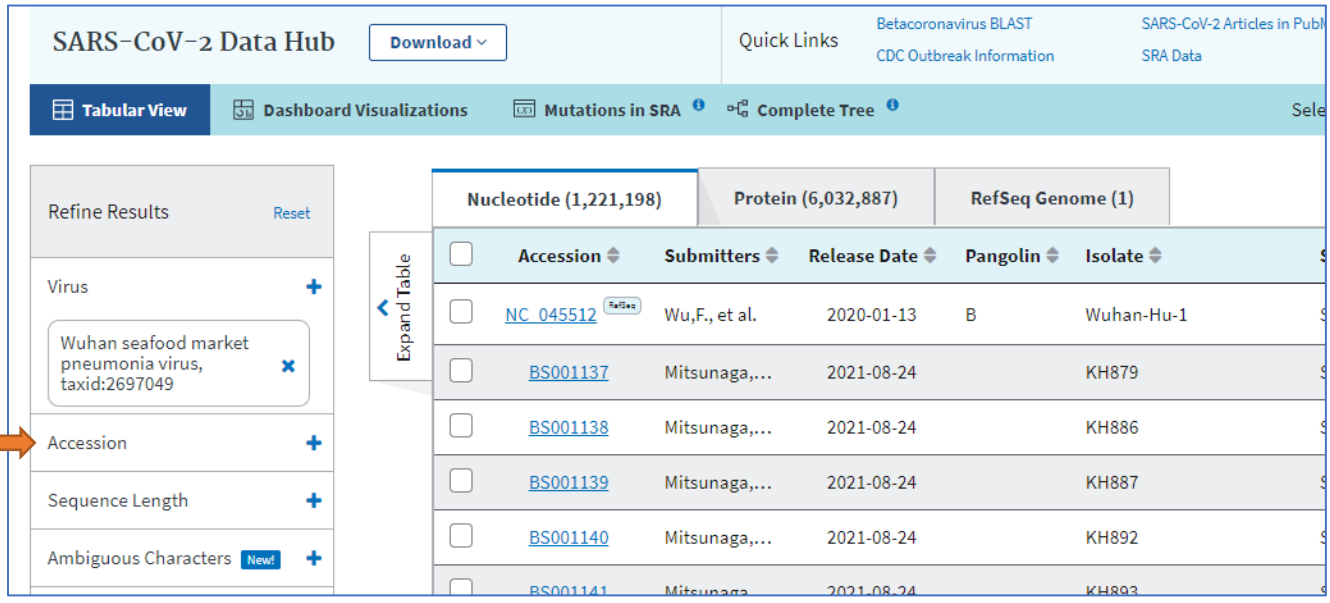
Navigating NCBI Using the NCBI Accession Number

Locating the Pangolin Lineage

1- You do not need a user account to follow these instructions. Navigate to the NCBI database at: https://www.ncbi.nlm.nih.gov/search/. Under COVID-19 resources click on **NCBI Virus**.

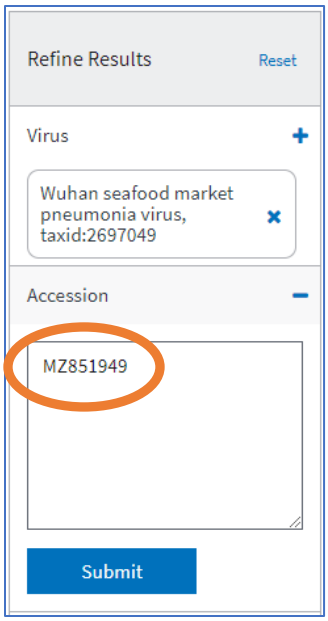


2- You will see the SARS-CoV-2 Data Hub. On the left-hand menu, under **Refine Results**, click **Accession**.

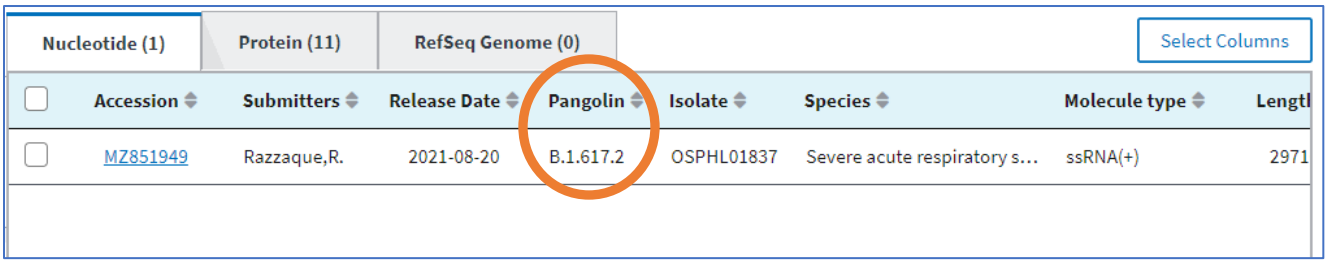


Created 8/24/2021

3- This will expand the pane > Enter in the **ACCESSION NUMBER** provided on the result report > then click **Submit**.



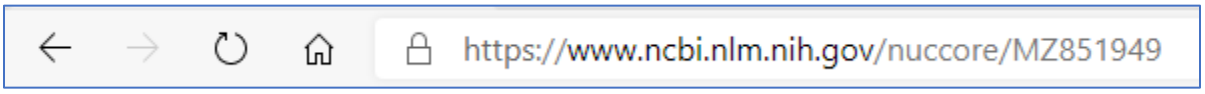
4- Result for the Accession number will show up on the table to the right. If you are only looking for the lineage, look under the Pangolin heading. Your use of NCBI is complete.



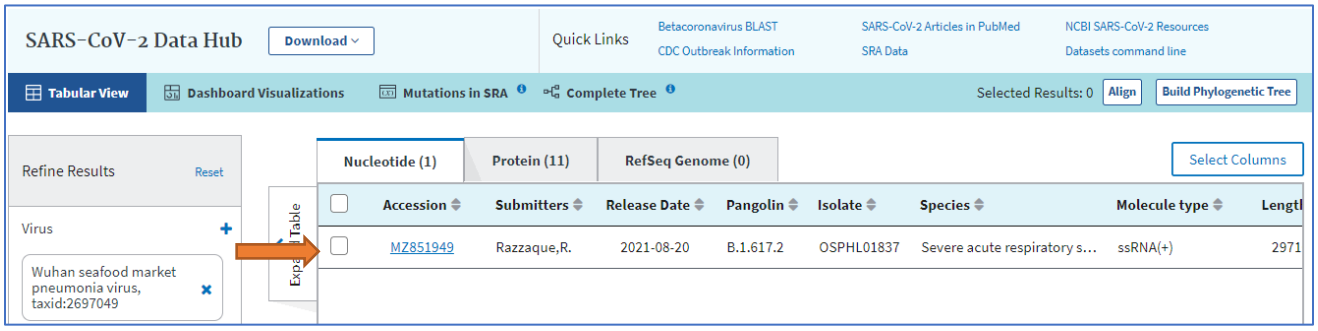
Locating the FASTA Sequence

Option #1: If only FASTA sequence is desired, you can obtain this quickly by entering the following into your browser: **https://www.ncbi.nlm.nih.gov/nuccore/[AccessionNumber]**

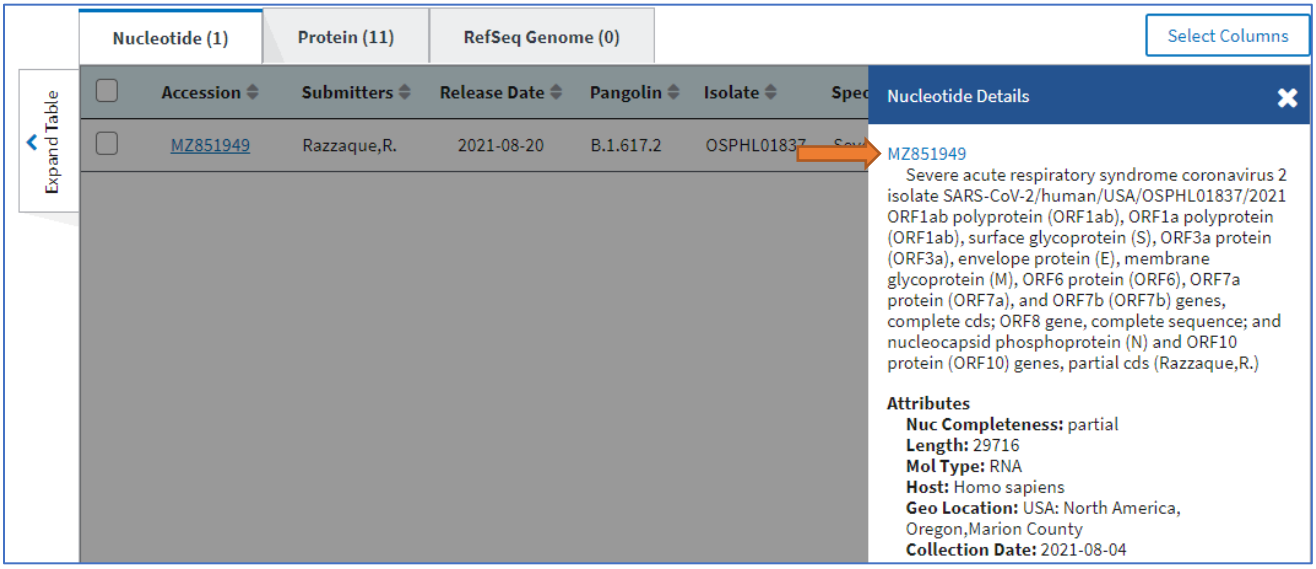
Here is an example: https://www.ncbi.nlm.nih.gov/nuccore/MZ851949



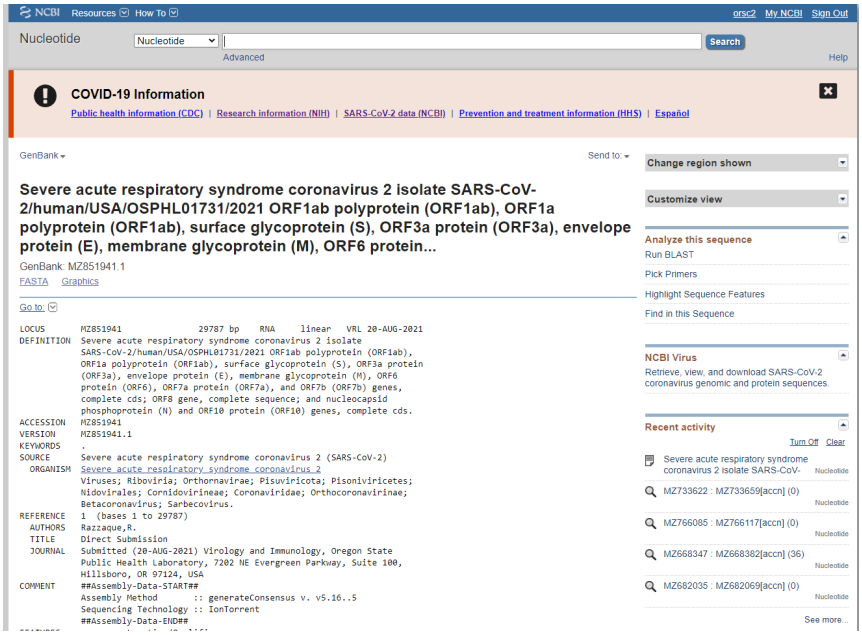
Option #2: If you have already navigated to the screen shown below, click on the **Accession number**.



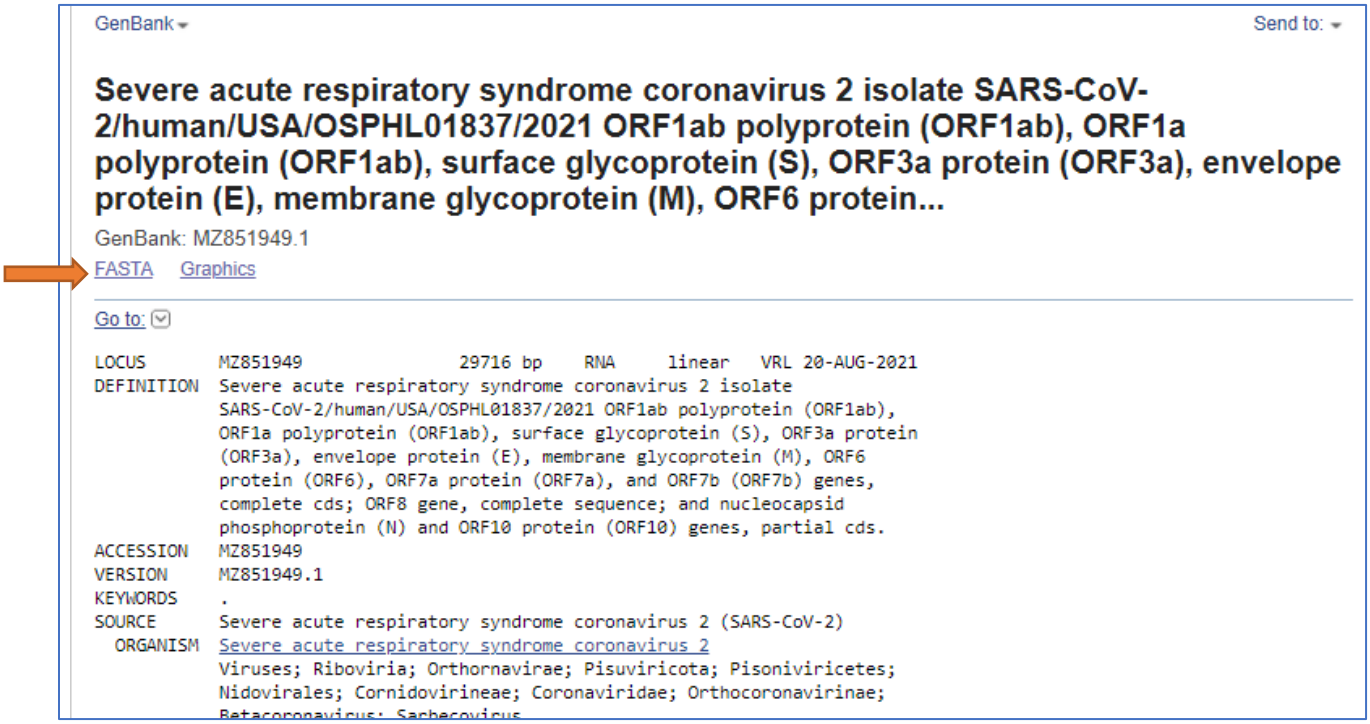
5- A window will open to show more information about the sequence. Click on the Accession Number.



6- A new browser tab will open showing all the details and sequence information.



7- Click on **FASTA** to get the FASTA sequence to be used for further analysis through other platforms.



8- FASTA sequence will open in this format:

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/OSPHL01837/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab), surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane glycoprotein (M), ORF6 protein...

GenBank: MZ851949.1

GenBank Graphics

```
>MZ851949.1 Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-
2/human/USA/OSPHL01837/2021 ORF1ab polyprotein (ORF1ab), ORF1a polyprotein (ORF1ab),
surface glycoprotein (S), ORF3a protein (ORF3a), envelope protein (E), membrane
glycoprotein (M), ORF6 protein (ORF6), ORF7a protein (ORF7a), and ORF7b (ORF7b) genes,
complete cds; ORF8 gene, complete sequence; and nucleocapsid phosphoprotein (N) and ORF10
protein (ORF10) genes, partial cds
GTAGATCTGTTCTCTAAACGAACTTTAAAATCTGTGTGGCTGTCACTCGGCTGCATGCTTAGTGCACTCA
CGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTG
CTTACGGTTTCGTCCGTTTTGCAGCCGATCATCAGCACATCTAGGTTTTGTCCGGGTGTGACCGAAAGGT
AAGATGGAGAGCCTTGTCCCTGGTTTCAACGAGAAAACACACGTCCAACTCAGTTTGCCTGTTTTACAGG
TTCGCGACGTGCTCGTACGTGGCTTTGGAGACTCCGTGGAGGAGGTCTTATCAGAGGCACGTCAACATCT
TAAAGATGGCACTTGTGGCTTAGTAGAAGTTGAAAAAGGCGTTTTGCCTCAACTTGAACAGCCCTATGTG
TTCATCAAACGTTCGGATGCTCGAACTGCACCTCATGGTCATGTTATGGTTGAGCTGGTAGCAGAACTCG
AAGGCATTCAGTACGGTCGTAGTGGTGAGACACTTGGTGTCCTTGTCCCTCATGTGGGCGAAATACCAGT
GGCTTACCGCAAGGTTCTTCTTCGTAAGAACGGTAATAAAGGAGCTGGTGGCCATAGTTACGGCGCCGAT
CTAAAGTCATTTGACTTAGGCGACGAGCTTGGCACTGATCCTTATGAAGATTTTCAAGAAAACTGGAACA
CTAAACATAGCAGTGGTGTTACCCGTGAACTCATGCGTGAGCTTAACGGAGGGGCATACACTCGCTATGT
CGATAACAACTTCTGTGGCCCTGATGGCTACCCTCTTGAGTGCATTAAAGACCTTCTAGCACGTGCTGGT
AAAGCTTCATGCACTTTGTCCGAACAACTGGACTTTATTGACACTAAGAGGGGTGTATACTGCTGCCGTG
AACATGAGCATGAAATTGCTTGGTACACGGAACGTTCTGAAAAGAGCTATGAATTGCAGACACCTTTTGA
AATTAAATTGGCAAAGAAATTTGACACCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNTTGAAAAGAAAAAGCTTGATGGCTTTATGGGTAGAATTCGATCTG
TCTATCCAGTTGCGTCACCAAATGAATGCAACCAAATGTGCCTTTCAACTCTCATGAAGTGTGATCATTG
TGGTGAAACTTCATGGCAGACGGGCGATTTTGTTAAAGCCACTTGCGAATTTTGTGGCACTGAGAATTTG
ACTAAAGAAGGTGCCACTACTTGTGGTTACTTACCCCAAAATGCTGTTGTTAAAATTTATTGTCCAGCAT
GTCACAATTCAGAAGTAGGACCTGAGCATAGTCTTGCCGAATACCATAATGAATCTGGCTTGAAAACCAT
TCTTCGTAAGGGTGGTCGCATTATTGCCTTTGGAGGCTGTGTGTTCTCTTATGTTGGTTGCCATAACAAG
TGTGCCTATTGGGTTCCACGTGCTAGCGCTAACATAGGTTGTAACCATACAGGTGTTGTTGGAGAAGGTT
CCGAAGGTCTTAATGACAACCTTCTTGAAATACTCCAAAAAGAGAAAGTCAACATCAATATTGTTGGTGA
CTTTAAACTTAATGAAGAGATCGCCATTATTTTGGCATCTTTTTCTGCTTCCACAAGTGCTTTTGTGGAA
ACTGTGAAAGGTTTGGATTATAAAGCATTCAAACAAATTGTTGAATCCTGTGGTAATTTTAAAGTTACAA
AAGGAAAAGCTAAAAAAGGTGCCTGGAATATTGGTGAACAGAAATCAATACTGAGTCCTCTTTATGCATT
TGCATCAGAGGCTGCTCGTGTTGTACGATCAATTTTCTCCCGCACTCTTGAAACTGCTCAAAATTCTGTG
CGTGTTTTACAGAAGGCCGCTATAACAATACTAGATGGAATTTCACAGTATTCACTGAGACTCATTGATG
CTATGATGTTCACATCTGATTTGGCTACTAACAATCTAGTTGTAATGGCCTACATTACAGGTGGTGTTGT
TCAGTTGACTTCGCAGTGGCTAACTAACATCTTTGGCACTGTTTATGAAAAACTCAAACCCGTCCTTGAT
TGGCTTGAAGAGAAGTTTAAGGAAGGTGTAGAGTTTCTTAGAGACGGTTGGGAAATTGTTAAATTTATCT
CAACCTGTGCTTGTGAAATTGTCGGTGGACAAATTGTCACCTGTGCAAAGGAAATTAAGGAGAGTGTTCA
GACATTCTTTAAGCTTGTAAATAAATTTTTGGCTTTGTGTGCTGACTCTATCATTATTGNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCACTCAAAGGGATTGTACAGAAAGTGTGTTAAAT
CCAGAGAAGAAACTGGCCTACTCATGCCTCTAAAAGCCCCAAAAGAAATTATCTTCTTAGAGGGAGAAAC
ACTTCCCACAGAAGTGTTAACAGAGGAAGTTGTCTTGAAAACTGGTGATTTACAACCATTAGAACAACCT
ACTAGTGAAGCTGTTGAAGCTCCATTGGTTGGTACACCAGTTTGTATTAACGGGCTTATGTTGCTCGAAA
TCAAAGACACAGAAAAGTACTGTGCCCTTGCACCTAATATGATGGTAACAAACAATACCTTCACACTCAA
AGGCGGTGCACCAACAAAGGTTACTTTTGGTGATGACACTGTGATAGAAGTGCAAGGTTACAAGAGTGTG
AATATCACTTTTGAACTTGATGAAAGGATTGATAAAGTACTTAATGAGAAGTGCTCTGCCTATACAGTTG
AACTCGGTACAGAAGTAAATGAGTTCGCCTGTGTTGTGGCAGATGCTGTCATAAAAACTTTGCAACCAGT
```