⊕ Notebook-INTRO_BIO-INFORMATICS_1_Using sequences

⸙ Forked from Notebook-INTRO_BIO-INFORMATICS_1_Using sequences

[1],[2]

[1]intro to bioinformatics lab notebook assignment; [2]UCSC

1 *Works for me*   This document is published without a DOI.

UCSC BME 22L

DOCUMENT CITATION

, 2020. Notebook-INTRO_BIO-INFORMATICS_1_Using sequences. **protocols.io**
https://protocols.io/view/notebook-intro-bio-informatics-1-using-sequences-bnsqmedw

FORK FROM

Forked from Notebook-INTRO_BIO-INFORMATICS_1_Using sequences,

LICENSE

This is an open access document distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

CREATED

Oct 22, 2020

LAST MODIFIED

Oct 22, 2020

DOCUMENT INTEGER ID

43568

DISCLAIMER:

INTRO BIO-INFORMATICS 1 : Using sequences

You are working with the local municipality health department to screen your community for COVID-19 cases. An effective readout for community-level COVID-19 infection rates is assessing the viral load ('titer', abundance, content) in community waste-water (Nature, CDC) and enables you to get a sense of how much virus is circulating in the community without a lot of individual testing -- also known as 'pooled testing'.

In order to detect viral presence, the health dept. is performing RNA sequencing on the waste-water entering the water treatment facility. They are detecting a subset of sequences that need characterization. Understanding the origin, nature, and implications of these RNA sequences will help the community assess its exposure and response to the pandemic.

The health department is providing you with the top 5 most abundant sequences from their screen in **FASTA** format:

```
#>header
#sequence

>seq_1
GGTTTTCGGATCGTAAAGCTCTGTTGTTAGGGAAGAACAAGTACCGTTCGAATAGGGCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCCCG
GCTCAACCGGGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAAC
ACCAGTGGCGAAGGCGA

>seq_2
TTGTGGCCATGGCGGCTGTGCCCATGGTGCTCAGTGCCATGGGCTTCACTGCGGCGGGAATCGCCTCGTCCTCCATAGCAGCCAAGATGATGTCCGCGGCG
GCCATTGCCAATGGGGGTGGAGTTGCCTCGGGCAGCCTTGTGGCTACTCTGCAGTCACTGG

>seq_3
ATGCACAAGGAGGAACATGAGGTGGCTGTGCTGGGGCCACCCCCCAGCACCATCCTTCCAAGGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCC
CGACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGTCTGGGC
TTCATAGCATTCGCCTACTCCGTGAAG

>seq_4
ATGTTTCATCTCGTTGACTTTCAGGTTACTATAGCAGAGATATTACTAATTATTATGAGGACTTTTAAAGTTTCCATTTGGAATCTTGATTACATCATAA
ACCTCATAATTAAAAATTTATCTAAGTCACTAACTGAGAATAAATATTCTCAATTAGAT
GAAGAGCAACCAATGGAGATTGATTAA
```

while the sequences have been given to you all in one file, it is best to assume that they are not derived from the same organism.

The health dept. believes that these sequences can help determine the **presence of Sars-CoV-2 and detect some signs of human response to infection**. Additionally, they'd like you to **confirm that these samples are derived from waste-water** as a control for future large scale experiments.

Please return a report that describes the **objectives** of your analysis, the **methods** you employed to generate each conclusion, and the **results** of those methods.
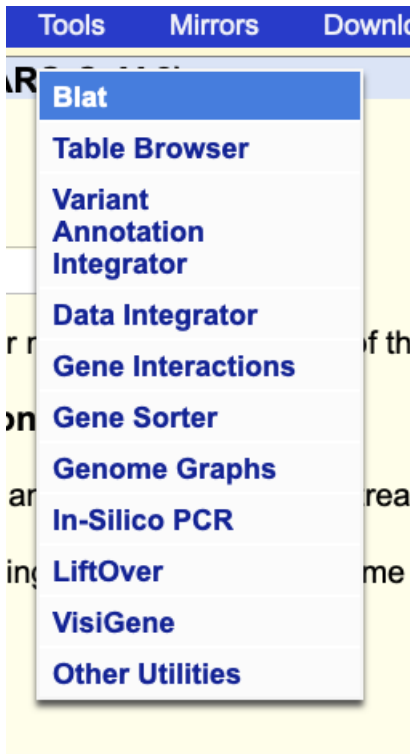
Results:

Please also provide the sequences in **FASTA** format with **headers re-named** to reflect the identities you've assigned each sequence.

For any **human gene** sequences you identify, please report the **GENCODE v32** *Biochemical and Signaling Pathways* it is associated with and whether or not we could study the gene using **mouse** as a **model organism**.

For any **Sars-CoV-2 gene** sequences you identify, use the Sars-CoV-2 Genome Browser's **BLAT** tool to see where it aligns to the genome, identify the gene it is derived from using the **Genes and Gene Predictions: NCBI Genes** track.

Where it would be helpful, attach a screenshot of any relevant output from BLAST/Genome Browser.

getting to the BLAT tool

setting the search to use SARS-CoV-2

Conclusion:

Finally, address the the interests of the health department in your conclusion:
(please cite the results from your analysis that support your conclusions)
1. **Can we confirm the presence of Sars-Cov-2?**
2. **Can we detect signs of human response to infection?**
3. **Can we confirm that these samples are likely derived from waste-water?**

OBJECTIVES
▪

METHODS

RESULTS

>
GGTTTTCGGATCGTAAAGCTCTGTTGTTAGGGAAGAACAAGTACCGTTCGAATAGGGCGGTACCTTGACGGTACCTAACCAGAAAGCCACGGCTAACTACG
TGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTGTCCGGAATTATTGGGCGTAAAGGGCTCGCAGGCGGTTTCTTAAGTCTGATGTGAAAGCCCCCG
GCTCAACCGGGGAGGGTCATTGGAAACTGGGGAACTTGAGTGCAGAAGAGGAGAGTGGAATTCCACGTGTAGCGGTGAAATGCGTAGAGATGTGGAGGAAC
ACCAGTGGCGAAGGCGA

>
TTGTGGCCATGGCGGCTGTGCCCATGGTGCTCAGTGCCATGGGCTTCACTGCGGCGGGAATCGCCTCGTCCTCCATAGCAGCCAAGATGATGTCCGCGGCG
GCCATTGCCAATGGGGGTGGAGTTGCCTCGGGCAGCCTTGTGGCTACTCTGCAGTCACTGG

>
ATGCACAAGGAGGAACATGAGGTGGCTGTGCTGGGGCCACCCCCCAGCACCATCCTTCCAAGGTCCACCGTGATCAACATCCACAGCGAGACCTCCGTGCC
CGACCATGTCGTCTGGTCCCTGTTCAACACCCTCTTCTTGAACTGGTGCTGTCTGGGC
TTCATAGCATTCGCCTACTCCGTGAAG

>
ATGTTTCATCTCGTTGACTTTCAGGTTACTATAGCAGAGATATTACTAATTATTATGAGGACTTTTAAAGTTTCCATTTGGAATCTTGATTACATCATAA
ACCTCATAATTAAAAATTTATCTAAGTCACTAACTGAGAATAAATATTCTCAATTAGAT
GAAGAGCAACCAATGGAGATTGATTAA

CONCLUSION