

Annotation Exercise

It will never really be this easy, but eventually you will be able to annotate your own data.

Where we are

- 13:30-14:00 – Primer Design to Amplify Microbial Genomes for Sequencing
- 14:00-14:15 – Primer Design Exercise
- 14:15-14:45 – Molecular Barcoding to Allow Multiplexed NGS
- 14:45-15:15 – Processing NGS Data – de novo and mapping assembly
- 15:15-15:30 – Break
- 15:30-15:45 – Assembly Exercise
- 15:45-16:15 – Annotation
- **16:15-16:30 – Annotation Exercise**
- 16:30-17:00 – Submitting Data to GenBank

Use our assembled genome sequence

- Try annotating it at VIGOR website
<http://www.jcvi.org/vigor/submission.php>
- Look at the .tbl file

- About VIGOR
- Run VIGOR
- Instructions & Test Samples
- Coronavirus
- Influenza virus
- Measles & mumps virus
- Metapneumovirus
- Norovirus
- Parainfluenza & Sendai virus
- Respiratory syncytial virus (RSV)
- Rhinovirus
- Rotavirus
- Rubella virus
- SARS
- Venezuelan equine encephalitis virus & alphavirus
- West Nile virus & Dengue

VIGOR - VIRAL GENOME ORF READER

Paste your sequences in **FASTA** format into the box below or upload **FASTA** file to the right

Browse...

No file selected.

```

CGGGTGAACAACCTTCTTTCTTCAGGTCCTACTATGCCCGGATGCAGCGGGTATCCCAACA
TGGATTTGGATTGCCTACTCCCCAGGAGTGGGTGCAGCACTTCTACCAAGAGGCAGCTC
CAGCACAATCTGATGTGGCTCTATTGAGATTTGTGAATCCAGACACGGGTAGGGTCTCTGT
TTGAGTGCAAACCTCATAAATCAGGCTATGTCACAGTGGCTCACACCGGTCAGCATGATT
TGGTCATCCCCCCAATGGCTATTTTAGGTTTGATTCTGGGTTAATCAGTTCTACACGC
TTGCCCCCATGGGAAACGGAACGGGGCGTAGGCGCGCCTTATAATGGCTGGAGCTTTCTT
TGCTGGATTGGCATCTGATGCCTTGGCTCTGGACTTGGTTCCCTAATCAATGCTGGGGC
TGGGGCCATCAACAAAAGATTGATTTGAAAATAATAGAAAATTGCAGCAAGCTTCCTT
CCAGTTTAGCAGTAATCTACAACAGGCTTCCTTTCAACACGATAAAGAGATGCTCCAAGC
ACAAATTGAGGCCACTAAAAAGTTGCAACAGGAAATGATGAAAGTCAAGCAGGCAATGCT
CCTAGAAGGTGGATTCTCTGAAACAGATGCAGCCCCTGGGGCAATCAACGCCCCCATGAC
AAAGTTTTGGACTGGAGCGGAACAAGGTAAGTGGGCCCTGATGCTAGGACCACAACATA
CAATGCAGGCCGCTTTTCCACCCCTCAACCTTCGGGGACGCTGCCAGGAAGAATCAATCC
CAGGACTCCTACCCCGCTCGGGGCTCCTCCAACACATCTTCTAATGCTTCCACTGCTAC
TTCTATACATTCAAATCAAACACTGTTTCAACGAGACTTGGTTCTACAGCTGTTCTGGTAC
CAATGTCTCGAGTCTCCCGTCAACTGCAAGGACTAGGAGTTGGGTTGAGGATCAAAACAG
AAATTTGTCACCTTTCATGAGGGGGGCTCATAACATATCGTTTGTACCCCACCATCTAG
CAGATCCTCTAGCCAAGGCACAGTCTCAACCGTGCCTAAAGAAGTTTTGGACTCCTGGAC
TGGCGCTTTC AACACGCGCAGGCAGCCTCTCTTCGCTCACATTCTAGGCGAGGGGAGTC
ACGGGTGTAA

```

Email address:

Select virus type:

[Report a Bug](#)



VIGOR - Norovirus

- About VIGOR
- Run VIGOR
- Instructions & Test Samples
- Coronavirus
- Influenza virus
- Measles & mumps viruses
- Metapneumovirus
- Norovirus
- Parainfluenza & Sendai viruses
- Respiratory syncytial virus (RSV)
- Rhinovirus
- Rotavirus
- Rubella virus

VIGOR SUBMISSION RECEIVED

Thank you for your submission.
 Your submission ID is '**dmzweb2_1898_1_WedAug211817402013**'.
 A notification email has been sent to '**tstockwe@jcvi.org**'
 You will receive another notification email once your job has been processed.
[Back to submission portal](#)
 Your results will be available at the following URL:
http://www.jcvi.org/vigor/results/dmzweb2_1898_1_WedAug211817402013/results.html



VIGOR - Norovirus

JCVI Home VIGOR Home

- About VIGOR
- Run VIGOR
- Instructions
- Coronavirus
- SARS
- Influenza virus
- Rhinovirus
- Rubella virus
- Metapneumovirus
- Norovirus
- Measles & mumps viruses
- Yellowfever & Japanese encephalitis viruses
- Respiratory syncytial virus (RSV)

VIGOR PREDICTION RESULTS

Your prediction results are ready for 'Norovirus'.

Your input FASTA file [view](#)
 The ORF prediction file [view](#)
 The CDS prediction file [view](#)
 The clustalw multiple alignment file [view](#)
 The GenBank TBL file [view](#)
 Download the entire set of files [here](#)
 How to download tarball click [here](#).

The submission ID is 'dmzweb2_1898_1_WedAug211817402013'
 The job ID is 'dmzweb2_1898_1_WedAug211817402013'

```

>Features contig00001
3      5102      gene
                gene      ORF1
3      5102      CDS
                gene      ORF1
                gene_syn      ORF1
                product genome polyprotein
                protein_id      contig00001.1
>Features      contig00001.1
3      992      mat_peptide
                product mat_peptide protein p48
993     2090     mat_peptide
                product mat_peptide NTPase
2091    2627    mat_peptide
                product mat_peptide p22k_1
2628    3026    mat_peptide
                product mat_peptide viral genome-linked protein
3027    3569    mat_peptide
                product mat_peptide 3C-like protease
3570    5099    mat_peptide
                product mat_peptide RNA-directed RNA polymerase
5083    6701     gene
                gene      capsid protein
                note      pseudo-gene
6704    7510     gene
                gene      ORF3
6704    7510     CDS
                gene      ORF3
                gene_syn      ORF3
                product protein VP2
                protein_id      contig00001.3

```

```
>contig00001.1 ORF1 1699 Aa 3 5102 genome polyprotein
MKMASNDASAAAVANSNNDTAKSSSDKMFSNMAVTFKRALGARPKQPPPREIPQRPPRPP
TPELVKKIIPPPPNGEDEVVVSYSADKGVSGLPPELSTVRQPEETNTAFSVPPLNQRENRD
AKEPLTGTILEMWDGEIYHYGLYVERGLVGLGVHKPPAAISLAKVELTPLSLFWRPVYTPQ
YLISPDTLKKLHGETFPYTAFDNNCYAFCCWVLDLNDLNSWLSRRMIQRTTGFFRPYQDWNR
KPLPTMDDSKLKKVANIFLCALSSLFTRPIKDIIGKLRPLNIINILASCDWT FAGIVESL
ILLAEELFGVFWT PPDVSAMIAPLLGDFELQGPEDLVVELVPVVMGGIGLVLGFTKEKIGR
MLSSAASTLRACKDLGAYGLEILKLVKWFPPKKEEANELAMVRSIEDAVLDLEAIENNH
MTLLKDKDSLATYMRSLDLEEEKARKLSTKSASPDIVGTINALLARIAAARSLVHRAKE
ELSSRPRPVVLMISGRPGIGKTHLAREVAKRIAASLTGDQRVGLIPRNGVDHWDAYKGER
VVLWDDYGM SNP IHDALRLQELADTCPLTLNCDRIENKGVFSDSVIIITTNLANPAPLD
YVNFEACSRRIDFLVYAEAPEVEKAKRDFPGQPDMMWKNFSSDFSHIKLALAPQGGFDKN
GNTPHGKGVKMTLTTGSLIARASGLLHERLDE FELQGPALTTFNFRNKVLAFRQLAAEN
KYGLIDTMKVGRQLKDVKTMPDLKQALKNISIKKCQIVYSGCTYTLESDGKGNVKVDRIQ
STSVQTNNELAGALHHLRCARIRYYVKCVQEALYSIIQIAGAAFVTTTRIIKRVNIQDLWS
KPQVENTEEATNKDGC PKPKDDEEFVISSDDIKTEGKKGKNTGRGKKHTAFSSKGLSDE
EYDEYKRIREERNGRYSIEEYLQDRDKYEEVAIARATEEDFCEEEEAIRQRIFRPTRK
QRKEERASLGLVTGSEIRKRNPD DFKPKGKLWADDDRSVDYNEKLSFEAPPSIWSRIVNF
GSGWGFVWSPSLFITSTHVIPQGAKEFFGVPIKQIQVHKSGEFCRLRFPPKPIRTDVTGMI
LEEGAPEGTVVTLLIKSTGELMPLAARMGTHATMKIQGRTVGGQMGMLLTGSNAKSMDL
GTTPGDCGCPYIYKRGNDYVVIGVHTAAARGGNTVICATQGGEGEATLEGGDSKGTYCGA
PILPGSAPKLSTKTKFWRSSSTAPLPPGTYEPAYLGGKDP RVKGGPSLQQVMRDQLKPFT
EPRGKPPKPSVLEAAKTI INVLEQTI DPPEKWSFAQACASLDKTTSSGHPHMRKNDCW
NGESFTGKLADQASKANLMFEEGKNMTPVYTGALKDELVKTDKIYGKIKRLLWGSDLAT
MIRCARA FGLMDELKAHCVTLP IRVGMNMNEDGPIIFEKHSRYRYHYDADYSRWDSTQQ
RAVLAAALEIMVKFSSEPHLAQVVAEDLLSPSVVDVGDFTISINEGLPSGVPCTSQWNSI
AHWLLTLCALSEVTNLS PDI IQANSLFSFYGDDEIVSTDIKLDPEKLTAKLKEYGLKPTR
PDKTEGPLVISEDLDGLTFLRRTVTRDPAGWFGKLEQSSILRQMYWTRGPNHEDPSESMI
PHSQRP IQLMSLLGEEAALHGPTFYSKISKLVIAELKEGGMDFYVPRQEP MFRWMRFS DLS
TWEGRNLAPS FVNEDGVE*
```

```
>contig00001.1.1 3 992 mat_peptide protein p48
MKMASNDASAAAVANSNNDTAKSSSDKMFSNMAVTFKRALGARPKQPPPREIPQRPPRPP
TPELVKKIIPPPPNGEDEVVVSYSADKGVSGLPPELSTVRQPEETNTAFSVPPLNQRENRD
AKEPLTGTILEMWDGEIYHYGLYVERGLVGLGVHKPPAAISLAKVELTPLSLFWRPVYTPQ
```


Try your own favorite virus

- Get a complete unsegmented viral genome from GenBank – search “nucleotide” database for the virus you want with qualifier [organism] AND “complete genome”[title]
- Download the fasta record
- Try annotating it at VIGOR website
- Look at the .tbl file