Submitting Data to GenBank

Enabling others to use your hard work.

J. Craig Venter™
INSTITUTE

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**

J. Craig Venter™ INSTITUTE

Genbank Project Registration

- **Project Registration with BioProject**

- Need to register the project with NCBI before the project starts.
- It provides an organizational framework to access all data associated with a project
- Describes project scope, material, and objectives using a controlled vocabulary.

- **Project details**
  - **Project type:** Genome Sequencing
  - **Title:** Sequencing of Vaccine Preventable Disease Agents – Measles
  - ***Attributes:***
    - **Scope:** Multiisolate;
    - **Material:** Genome;
    - **Capture:** Whole;
  - **Description:** Detailed description of the project
  - **Relevance:** Medical
  - **Objectives**: Sequence, Assembly, Annotation

J. Craig Venter™ INSTITUTE

Minimum Data Type Required for Submission

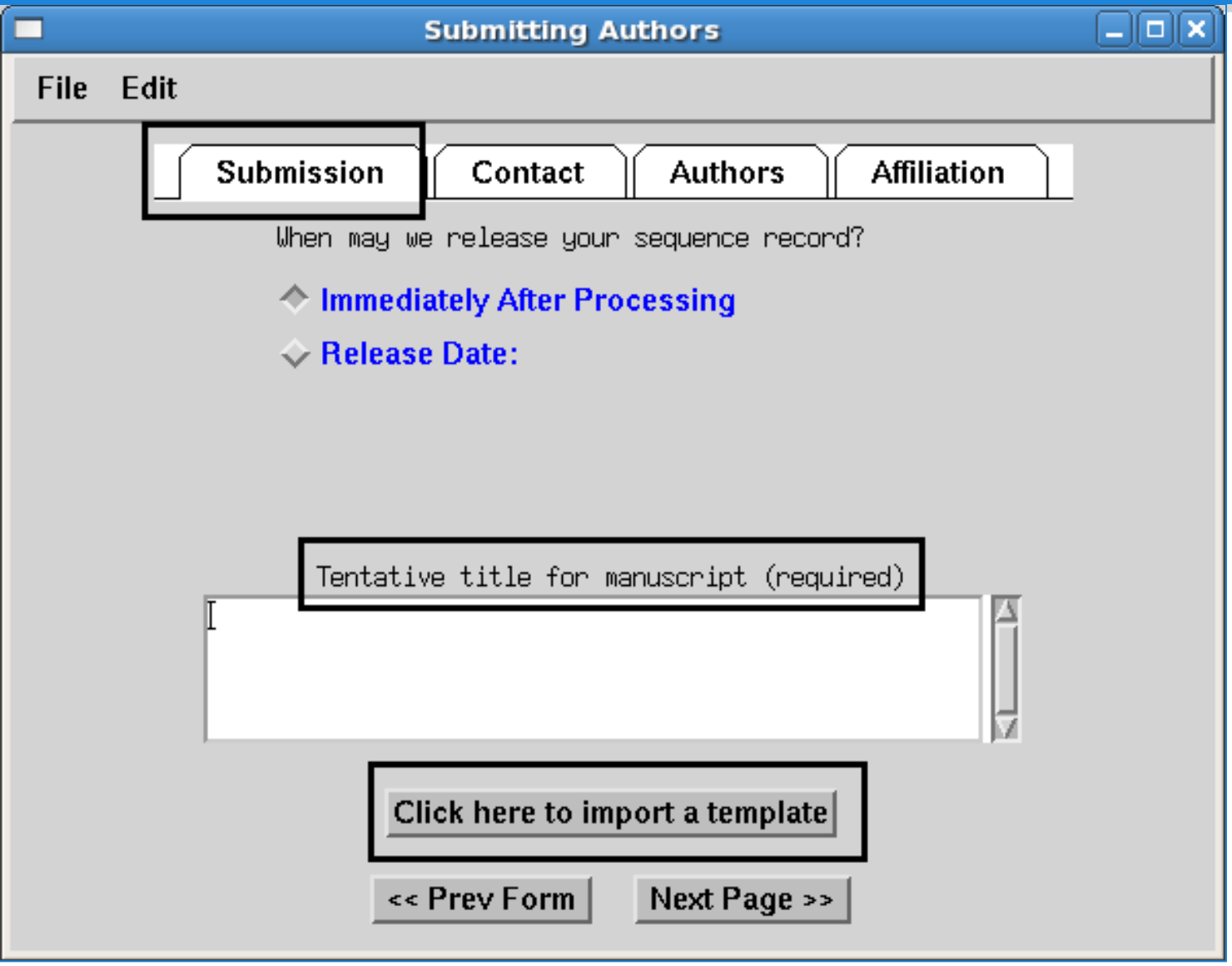
- **Here is a list of minimum metadata associated with a sample that needs to be submitted with the sequence**
- Organism name (it should be just the virus name without strain information in it.)
- Strain name
- Host from which the organism is isolated
- Collection date
- Country where isolate was collected from.
- Name, contact email and institution affiliation address of the Principal Investigator submitting the sequence.
- All these go in what the genbank calls the template file. (It can be generated using sequin)

J. Craig Venter™ INSTITUTE

Other Files Required for Submission

- **Sequence file in fasta format**
- **Asn file generated using Sequin OR tbl2asn**
- **If the sequence has gaps, submit an associated gapped fasta file with the estimated number of NNN denoting gaps OR with 100 NNNs in the gap region if the gap is of unknown size.**

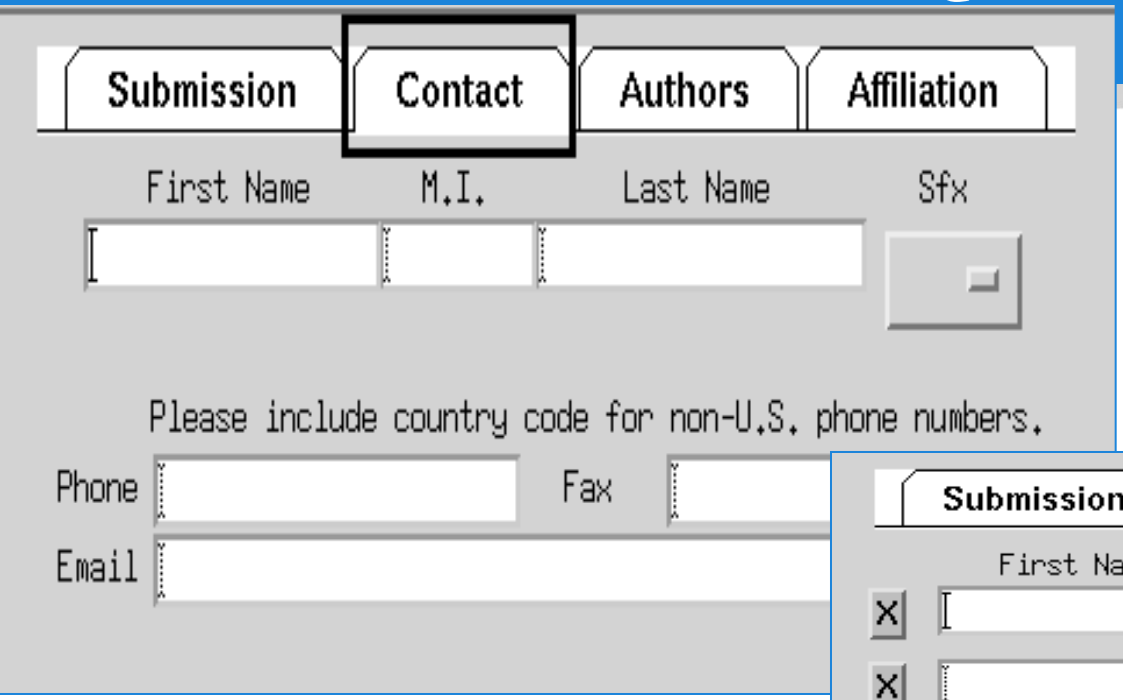
SEQUIN

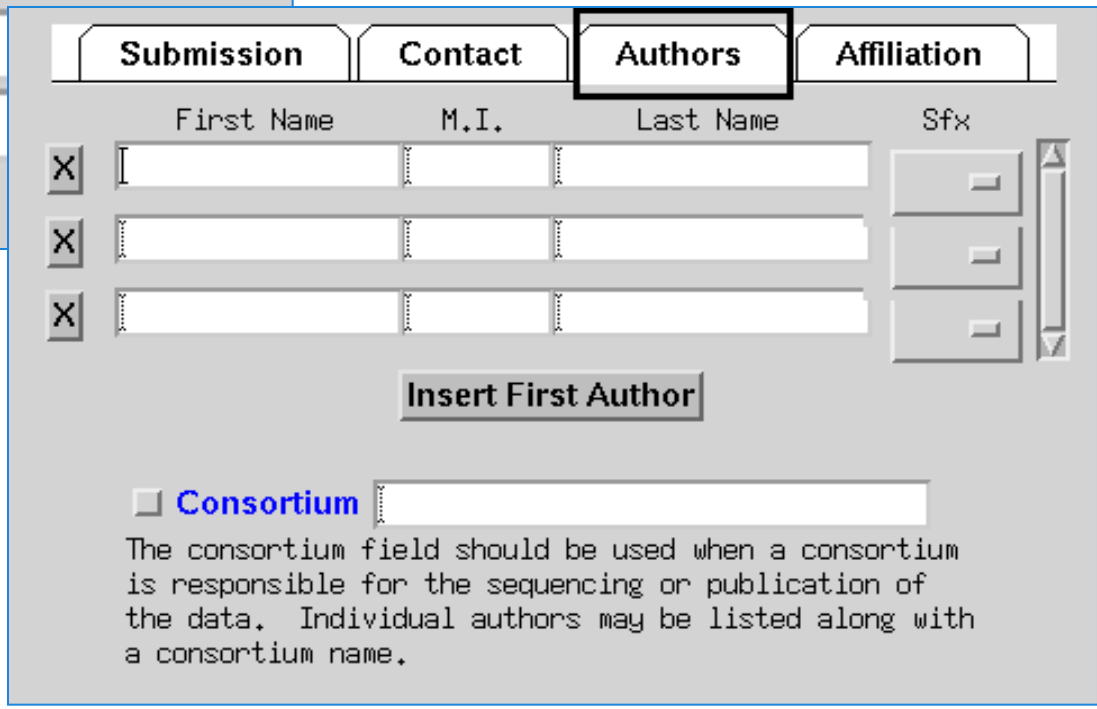


Sequin is a tool from NCBI that is used to generate files in the “asn” format which is the required format for submission of sequence to Genbank

J. Craig Venter INSTITUTE

SEQUIN



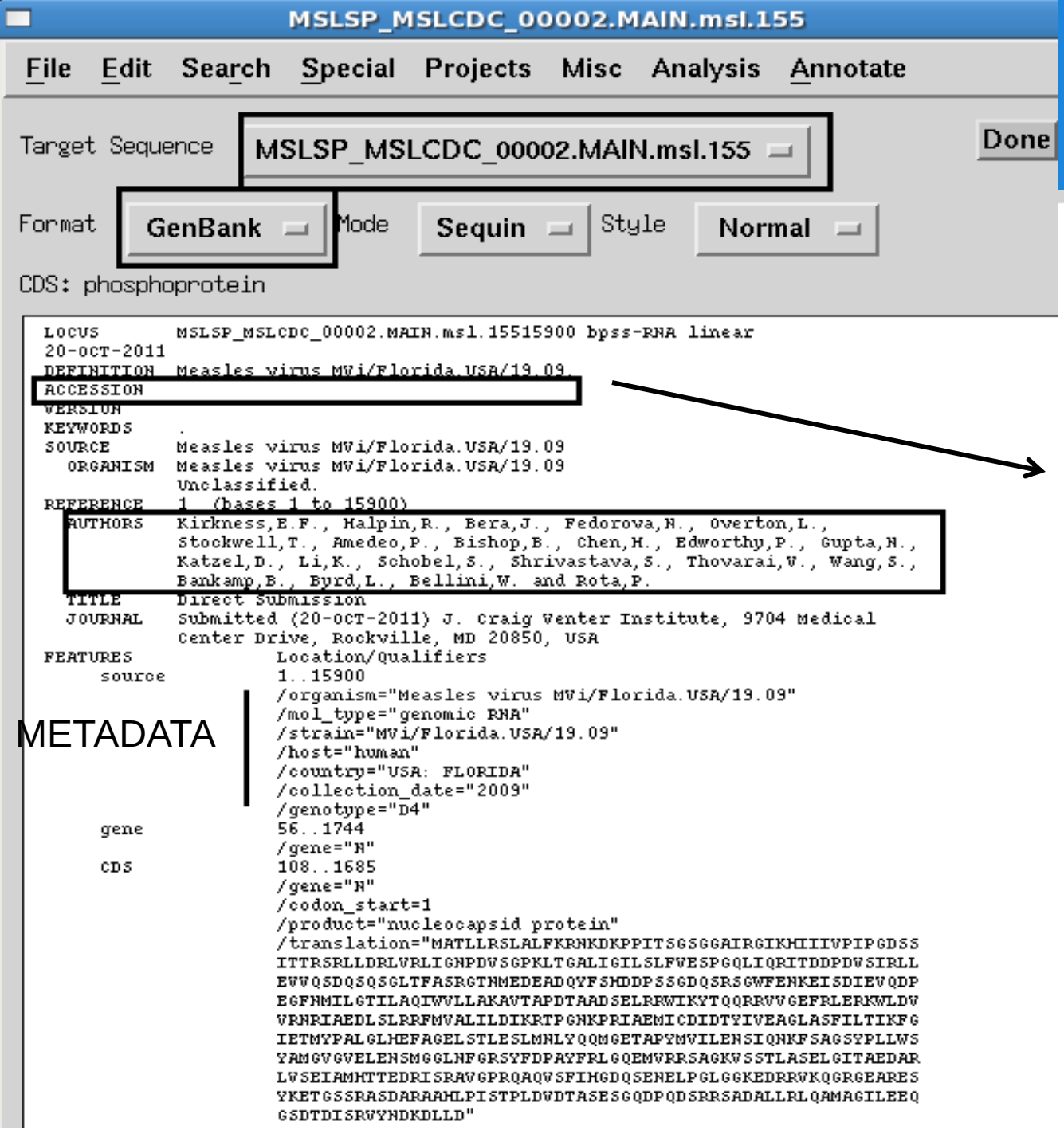


J. Craig Venter™ INSTITUTE

SEQUIN

- The template can be saved into a text file which can be reused; therefore not requiring to fill this multiple times.
- There is also a web version for submitting data to NCBI. Each data type has different requirements; links are in the resources.

J. Craig Venter INSTITUTE



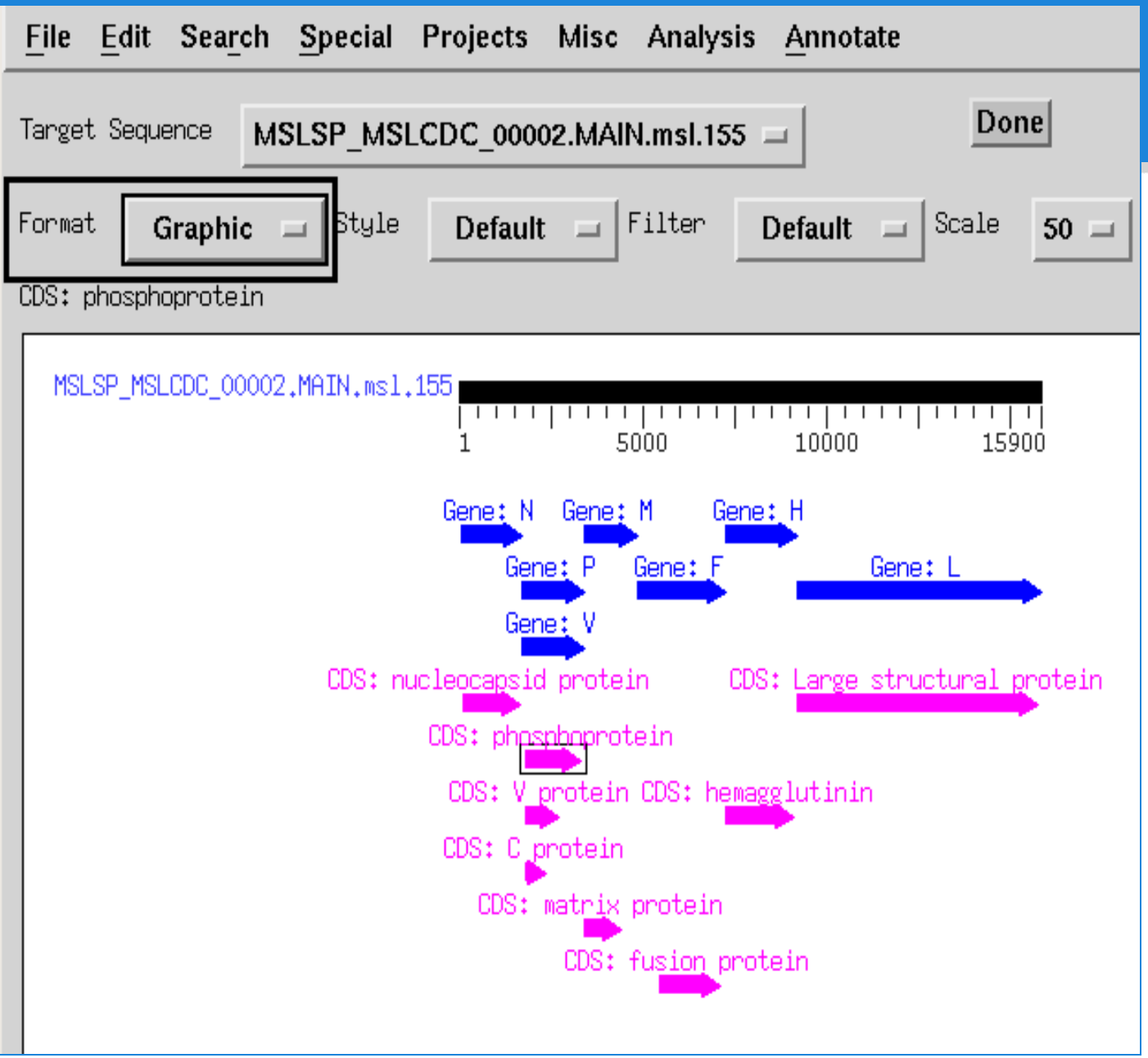
METADATA

Genbank Format

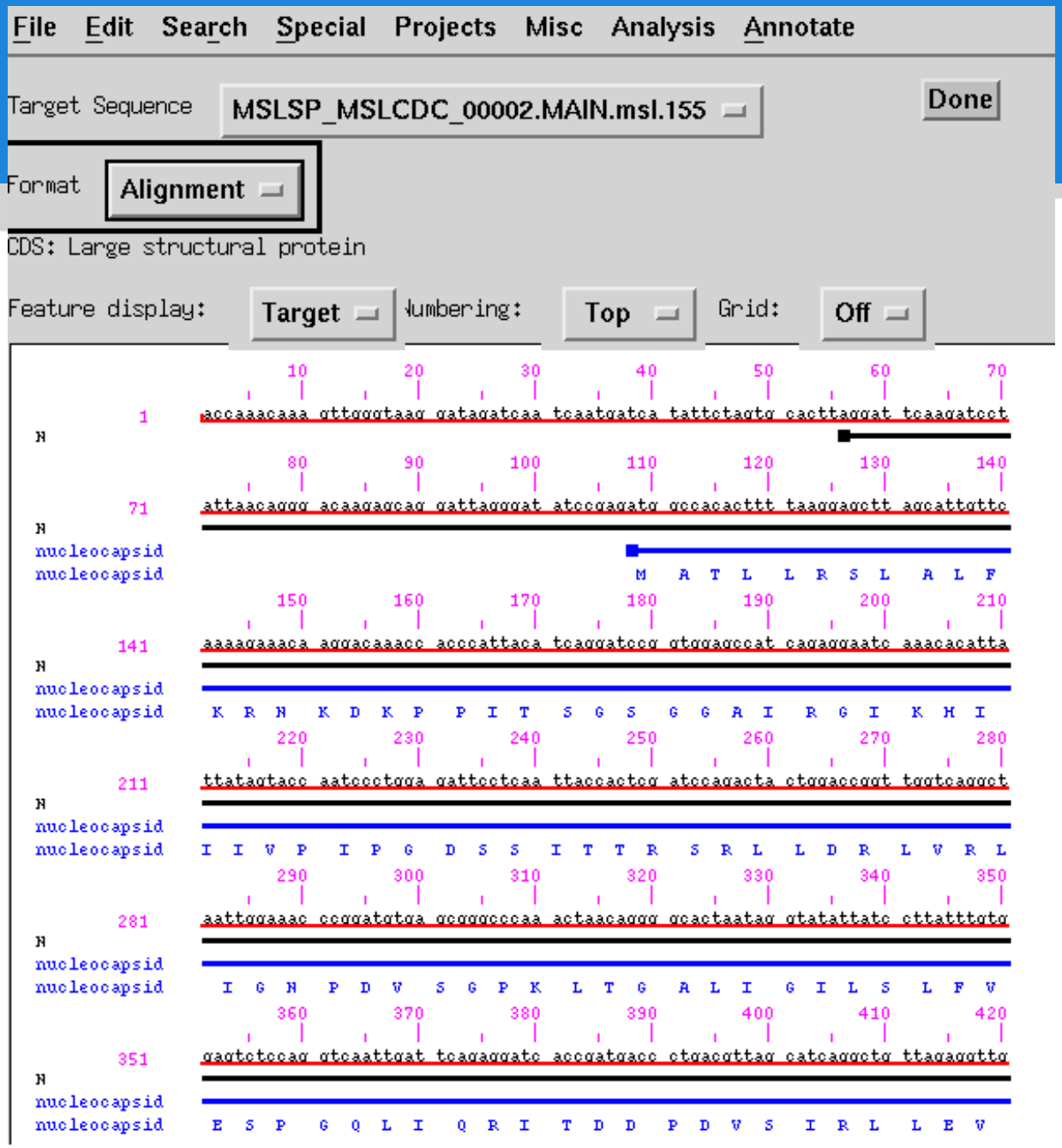
Accessions will be assigned when accepted by Genbank

J. Craig Venter INSTITUTE

Graphic Format



J. Craig Venter™ INSTITUTE



Alignment Format

Sequin is a very powerful tool for checking the accuracy of annotation

J. Craig Venter INSTITUTE

Resources for Genbank Submission

- **BioProject registration**

http://www.ncbi.nlm.nih.gov/bioproject/

- **Procedure for submitting data to Genbank**

http://www.ncbi.nlm.nih.gov/guide/howto/submit-data/

- **Bacterial Genome Submission Guide**

http://www.ncbi.nlm.nih.gov/genbank/genomesubmit

https://submit.ncbi.nlm.nih.gov/

- **Consistency Check and Validation tool for microbial genomes**

http://www.ncbi.nlm.nih.gov/genomes/frameshifts/frameshifts.cgi

- **Instruction for submitting Whole Genome Shotgun Sequences**

http://www.ncbi.nlm.nih.gov/genbank/wgs

- **Sequin**

http://www.ncbi.nlm.nih.gov/projects/Sequin/

J. Craig Venter™ INSTITUTE

You have been great for making it to the end!!

- Thanks for attending and staying!
- You won't remember everything you heard today… This workshop was to give you a "taste" of Bioinformatics for Viral Genomics.
- I like e-mail if you have questions later on tstockwell@jcvi.org
- Please consider our summer internships (applications due in March)

J. Craig Venter™ INSTITUTE

Questions?

J. Craig Venter™
INSTITUTE