

Script P6: HPyV Detection

HANNIGAN GD, GRICE EA, ET AL.

Abstract

This protocol provides a method for detecting Human Polyomaviruses (HPyVs) in our virome and whole metagenome datasets. For genome reference, download the HPyV genomes (in fasta format) from NCBI (nucleotide) using the search terms "Human Polyomavirus" AND "Complete Genome". This is used in the rest of the analyses. Based on the methods from the following publication:

Hannigan, Geoffrey D., et al. "The Human Skin Double-Stranded DNA Virome: Topographical and Temporal Diversity, Genetic Enrichment, and Dynamic Associations with the Host Microbiome." *mBio* 6.5 (2015): e01578-15.

Citation: HANNIGAN GD, GRICE EA, ET AL. Script P6: HPyV Detection. **protocols.io**

dx.doi.org/10.17504/protocols.io.efxbbpn

Published: 10 Mar 2016

Guidelines

Required Software:

- NCBI's BLAST+ v2.2.0

Relevant Files

- Output: HPyV_Detection

Before start

Perl scripts and other supplemental information available at:

https://figshare.com/articles/The_Human_Skin_dsDNA_Virome_Topographical_and_Temporal_Diversity_Genetic_Enrichment_and_Dynamic_Associations_with_the_Host_Microbiome/1281248

Protocol

Step 1.

Download the HPyV genome reference from NCBI (nucleotide) using the search terms "Human Polyomavirus" AND "Complete Genome".

 **LINK:**

<http://www.ncbi.nlm.nih.gov/refseq/>

Step 2.

Make an HPyV blast database using the fasta file from NCBI.

```
cmd COMMAND
makeblastdb -dbtype nucl -
in ./references/HPyV_reference_genomes/HPyV_complete_genomes_no_block.fasta -
out ./references/HPyV_reference_genomes/HPyV_database
```

Step 3.

Count the number of reads in the negative cleaned virome samples that are homologous to known HPyV genomes.

```
cmd COMMAND
mkdir ./detect_HPyV_in_dataset
mkdir ./detect_HPyV_in_dataset/negative_clean_virome_hits

run.blastn.hpyv () {
    blastn -query ./negative_clean_seqs/${1} -
out ./detect_HPyV_in_dataset/negative_clean_virome_hits/${1} -
db ./references/HPyV_reference_genomes/HPyV_database -outfmt 6 -num_threads 2 -
max_target_seqs 1 -evaluate 1e-3
}
export -f run.blastn.hpyv
ls ./negative_clean_seqs/* | sed -e 's/^.*\/*\/*\/*g' | xargs -I {} --max-procs=128 sh -
c 'run.blastn.hpyv {}'
wait
```

Step 4.

Count the number of reads in the non human decontaminated virome samples that are similar to known HPyV genomes.

```
cmd COMMAND
mkdir ./detect_HPyV_in_dataset/human_deconseq_virome_hits
mkdir ./detect_HPyV_in_dataset/human_deconseq_virome_fastas
```

Step 5.

Convert the fastq files to fasta before performing blastn.

```
cmd COMMAND
for file in $(ls ./cont_fastq/*R1* | sed -e 's/^.*\/*\/*\/*g' | sed 's/\.*fastq//'); do
    ./idba_ud-1.0.9/bin/fq2fa ./cont_fastq/${file}.fastq ./detect_HPyV_in_dataset/human_deconseq_virome_fastas/${file}.fasta
done
```

Step 6.

Run blastn.

 **SOFTWARE PACKAGE (Unix)**

BLAST Toolkit, 2.2.0 

NCBI
ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/

```
cmd COMMAND
mkdir ./detect_HPyV_in_dataset/errors
run.blastn.hpyv () {
    module load ncbi-blast-2.2.0
    blastn -query ./detect_HPyV_in_dataset/human_deconseq_virome_fastas/${1} -
out ./detect_HPyV_in_dataset/human_deconseq_virome_hits/${1} -
db ./references/HPyV_reference_genomes/HPyV_database -outfmt 6 -num_threads 2 -
max_target_seqs 1 -evaluate 1e-3
}
export -f run.blastn.hpyv
ls ./detect_HPyV_in_dataset/human_deconseq_virome_fastas/* | sed -
e 's/^.*\/*\/*\/*g' | xargs -I {} --max-procs=128 sh -c 'run.blastn.hpyv {}'
wait
```

Step 7.

Count the number of reads in the human decontaminated whole metagenome samples that are similar to known HPyV genomes.

cmd **COMMAND**

```
mkdir ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met
mkdir ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met_fastas
```

Step 8.

Convert the fastq files to fasta before performing blastn.

cmd **COMMAND**

```
for file in $(ls ./clean_phix_fastq/*R1* | sed -e 's/^.*\./.*\\.//g' | sed 's/\\.fastq//'); do
    ./idba_ud-1.0.9/bin/fq2fa ./clean_phix_fastq/${file}.fastq ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met_fastas/${file}.fasta
done
```

```
run.blastn.hpyv () {
    blastn -query ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met_fastas/${1} -
    out ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met/${1} -
    db ./references/HPyV_reference_genomes/HPyV_database -outfmt 6 -num_threads 2 -
    max_target_seqs 1 -evaluate 1e-3
}
export -f run.blastn.hpyv
ls ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met_fastas/* | sed -
e 's/^.*\./.*\\.//g' | xargs -I {} --max-procs=128 sh -c 'run.blastn.hpyv {}'
wait
```

Step 9.

Count the number of reads in the non human decontaminated whole metagenome samples that are similar to known HPyV genomes.

cmd **COMMAND**

```
mkdir ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met
mkdir ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met_fastas
```

Step 10.

Convert the fastq files to fasta before performing blastn.

cmd **COMMAND**

```
for file in $(ls ./deconseq_fastq/*_R1/*_cont* | sed -
e 's/^.*deconseq_fastq\\/(.*fastq\\).*/\\1/' | sed 's/\\.fastq//'); do
    echo Converting file ${file}...
    ./idba_ud-1.0.9/bin/fq2fa ./deconseq_fastq/${file}.fastq/*_cont.fq ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met_fastas/${file}.fasta
done
```

```
run.blastn.hpyv () {
    blastn -query ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met_fastas/${1} -
    out ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met/${1} -
    db ./references/HPyV_reference_genomes/HPyV_database -outfmt 6 -num_threads 2 -
    max_target_seqs 1 -evaluate 1e-3
}
export -f run.blastn.hpyv
ls ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met_fastas/* | sed -
e 's/^.*\./.*\\.//g' | xargs -I {} --max-procs=128 sh -c 'run.blastn.hpyv {}'
wait
```

Step 11.

Quantify the numbers of HPyV reads in all of the sample sets.

cmd **COMMAND**

```
WC -
l ./detect_HPyV_in_dataset/negative_clean_virome_hits/* | sed 's/^ *///g' | sed 's/ /\\t/' |
```

```

grep -
v 'total' | sed 's/\.\./*\///' | sed 's/_R1\.fa//' > ./detect_HPyV_in_dataset/negative_clean_viroome_hit_HPyV_count.tsv
wc -
l ./detect_HPyV_in_dataset/human_deconseq_viroome_hits/* | sed 's/^ *//g' | sed 's/ /\t/' |
grep -
v 'total' | sed 's/\.\./*\///' | sed 's/_R1\.fasta//' > ./detect_HPyV_in_dataset/human_contaminated_viroome_hit_HPyV_count.tsv
wc -
l ./detect_HPyV_in_dataset/human_and_phiX_decon_skin_met/* | sed 's/^ *//g' | sed 's/ /\t/' |
| grep -
v 'total' | sed 's/\.\./*\///' | sed 's/_R1\.fasta//' > ./detect_HPyV_in_dataset/decontaminated_skin_met_hit_HPyV_count.tsv
wc -
l ./detect_HPyV_in_dataset/human_and_phiX_cont_skin_met/* | sed 's/^ *//g' | sed 's/ /\t/' |
| grep -
v 'total' | sed 's/\.\./*\///' | sed 's/_R1\.fasta//' > ./detect_HPyV_in_dataset/human_contaminated_skin_met_hit_HPyV_count.tsv

```