

Overview of HPVViewer

Contents

Description	1
Installation	1
Parameters	1
Results	2
Basic Usage (demo)	2
Workflow	3

Description

HPVViewer is a tool for **genotyping and quantification of HPV from metagenomic or human genomic shotgun sequencing data**. We designed it to improve performance by masking nonspecific sequences from reference genomes and directly identifying HPV short DNA reads. It contains two HPV databases with different masking strategies, repeat-mask and homology-mask and one homology distance matrix to choose between those two databases.

If you use the HPVViewer software, please cite our manuscript:

Yuhan Hao, Liying Yang, Antonio Galvao Neto, Milan R. Amin, Dervla Kelly, Stuart M. Brown, Ryan C. Branski, Zhiheng Pei. **“HPVViewer: Sensitive and specific genotyping of human papillomavirus in metagenomic DNA”** (Submitted)

Installation

```
$ git clone https://github.com/yuhanH/HPVViewer.git
```

Pre-requisites

Python (2.7+)

Python packages (sys, getopt, subprocess)

Bowtie2: <http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>

SAMtools: <http://www.htslib.org/>

Bedtools: <http://bedtools.readthedocs.io/en/latest/>

Parameters

Required

- input files (-U or -1 -2): fastq files (or fastq.gz), unpaired (-U unpaired.fastq) or R1,R2 paired (-1 R1.fastq -2 R2.fastq)
- output file name (-o)

Optional

- a) database mask type (**-m**): hybrid-mask(default), repeat-mask, homology-mask. repeat-mask is a more sensitive mode; and homology-mask is suggested when some types of HPV are present in large abundance which may lead to false positive of other types of HPV.
- b) number of threaded used in bowtie2 alignment (**-p**)
- c) minimal coverage threshold to determine HPV present (**-cov**), default is 150 bp (1.5 x average length of your reads).

Results

- a) output_HPV_summary.txt has three columns with types of HPV present, number of reads per kilobase (RPK) for the matching HPV, and number of reads of the matching HPV.

$$\text{average coverage} = \frac{RPK * \text{average reads length}}{1000}$$

- b) alignment results after bowtie2: output.sam, output.bam

Basic Usage (demo)

```
python HPVviewer2.py -U test_unpaired.fastq -o TEST
```

```
more TEST/TEST_HPV_profile.txt
```

```
## HPV_type RPK(reads_per_kilobase) count_of_reads
## HPV5 114.3199 859
## HPV7 109.0264 860
## HPV21 120.3347 906
## HPV49 109.3479 820
## HPV66 123.1417 936
## HPV69 108.1513 812
## HPV73 117.2339 868
## HPV88 126.1261 924
## HPV101 115.1348 833
## HPV117 116.1049 899
```

Workflow

