

MHashGen

Version 1.4.2 04/15/2011 (Linux)

Usage

MHashGen

-h - Prints this message.

-v -PrintsMHashGen version.

MHashGen -d<DataDir> -f <HashFileName> -s<SeqFileName>

-d<DataDir> Use DataDir for MHashGen (MBlastX) internal files.

Default -use\$MBLASTX_DATADIR environment setting or current directory.

This argument overrides MBLASTX_DATADIR environment variable.

-f <HashFileName> -Name of Hash Files produced. Default 'ProtHash'

-s<SeqFileName> -Fasta file containing reference sequences.

-T <InputDatabaseType> , where <InputDatabaseType> is 'N' for Protein database; 'T' for Nucleotide database.

Usage note

- ⤴ *mhashgen_mtblastx*, with *-T T* option should be used to generate Reference data files from Nucleotide databases.
- ⤴ *mhashgen* with *-T N* option should be used to generate Reference data files from Protein databases.

Known Limitations

The executable *mhashgen* when run with a nucleotide database and/or with *-T T* option is known to produce unpredictable behavior.

The executable *mhashgen_mtblastx* when run with a protein database and/or with *-T N* or with no *-T* option is known to produce unpredictable behavior.

Example Usage

1. Specifying the Reference Sequences file in FASTA format

```
mhashgen -s nr.downloaded.091003.fna
```

The *-s* parameter specifies the Reference Sequence data file, which is expected to be a valid FASTA file. This parameter specification is mandatory and there is no default value. The Reference Sequence

file is expected to be in the current directory and the Hash Files for the reference database, prefixed 'ProtHash', are created in the current directory. In case the MBLASTX_DATADIR environment variable is set, the Reference Sequence file is expected to be in the directory pointed to by this environment variable and the hash files are created in the directory pointed to by this environment variable.

2. Specifying the data directory for MblastX

```
mhashgen -s nr.downloaded.091003.fna -d /home/mblastx/data
```

With the *-d* parameter, the Reference Sequence file is expected to be in the directory pointed to by this parameter and the hash files are created in the directory pointed to by this parameter. This is irrespective of the setting for the MBLASTX_DATADIR environment variable. In the above example, the data directory used is */home/mblastx/data*.

3. Specifying the Hash File name

```
mhashgen -s nr.downloaded.091003.fna -f nr_091003
```

With the `-f` parameters, the hash file prefix created using the Sequence Reference File to be used by the MBlastX is specified. In the above example, the file prefix used is 'nr_091003', instead of the default 'ProtHash'. This parameter is useful if more than one set of hash files are created in the same directory.