

Settings and Install

0. Requirements to install the HLA-HD

- **Console application**

HLA-HD only works on CUI based console system. Standard UNIX based PCs such as the Mac OS (terminal application) or the Linux OS installed it. If you use the Windows PC, install UNIX-like system such as [CYGWIN](#).

- **GNU C/C++ compiler**

The compiler is needed to build the HLA-HD.

- **Bowtie2**

Set its path to your environment variables

Bash Series:

```
export PATH=$PATH:path_to_bowtie2
```

Csh Series:

```
setenv PATH "${PATH}:"path_to_bowtie2"
```

Check the provider's website for installing and usage of other software.

1. Download compressed source file of HLAHD (*hlahd.x.y.z.tar.gz**) from download site

URL: <https://www.genome.med.kyoto-u.ac.jp/HLA-HD/download-request/>

The values of *x*, *y* and *z* depend on the version of HLA-HD.

2. Copy source file to appropriate location of your PC

Hereafter, we assume *hlahd.x.y.z.tar.gz* is copied in *target_path* directory.

3. Extract compressed source file

```
> cd target_path  
> tar -zxvf hlahd.x.y.z.tar.gz
```

4. Build commands by using *install.sh*

```
> cd hlahd.x.y.z  
> bash install.sh
```

5. Set path to installed commands

Add a following line to the setting file whose path is listed in **Table 1**,

Bash Series:

```
export PATH=$PATH:target_path/hlahd.x.y.z/bin
```

Csh Series:

```
setenv PATH "${PATH}:target_path/hlahd.x.y.z/bin"
```

6. Enable path to installed commands

Run `source` command as same as step 3 or re-login (logout and then login).

Example of execution

Here, we demonstrate steps in the execution of the HLA-HD for a sample (SRR794547) data from the '1000 Genomes Project', obtained from the 'HapMap Project'.

Assumptions

- HLAHD is installed on `~/local/hlahd.1.4.0`
- Target paired-end FASTQ files are in `$PWD/data` directory
- The login shell is `/bin/bash`

Execution

1. Confirm where hlahd.sh is located

```
$ which hlahd.sh
~/local/hlahd.1.4.0/bin/hlahd.sh
$
```

2. Confirm where target paired-end FASTQ files are located

```
$ ls data
SRR794547_1.fastq SRR794547_2.fastq
$
```

3. Execute hlahd.sh

Here, we assume following parameters are used in the execution,

- Number of threads in execution is 4
- Minimum length of reads is 100
- File paths to the paired-end FASTQs are `data/SRR794547_1.fastq` and `data/SRR794547_2.fastq`
- The default paths are used for `freq.` data directory, `HLA_gene.split.txt` and dictionary directory
- ID name is `SRR794547`
- Output directory is the current directory (= `.`).

```
$ hlahd.sh -t 4 -m 100 -f ~/local/hlahd.1.4.0/freq_data \  
> data/SRR794547_1.fastq data/SRR794547_2.fastq \  
> ~/local/hlahd.1.4.0/HLA_gene.split.txt \  
> ~/local/hlahd.1.4.0/dictionary SRR794547 .  
  
HLA-HD version 1.4.0  
22012539 reads; of these:  
  22012539 (100.00%) were unpaired; of these:  
    21843740 (99.23%) aligned 0 times  
    16546 (0.08%) aligned exactly 1 time  
    152253 (0.69%) aligned >1 times  
0.77% overall alignment rate  
22012539 reads; of these:  
  22012539 (100.00%) were unpaired; of these:  
    21840736 (99.22%) aligned 0 times  
    16599 (0.08%) aligned exactly 1 time  
    155204 (0.71%) aligned >1 times  
0.78% overall alignment rate  
. . .  
171803 reads; of these:  
  171803 (100.00%) were unpaired; of these:  
    8641 (5.03%) aligned 0 times  
    15008 (8.74%) aligned exactly 1 time  
    148154 (86.23%) aligned >1 times  
94.97% overall alignment rate  
$
```

4. Check whether the hlahd.sh is executed correctly or not

```
$ ls -F SRR794547
estimation.sh exon/ intron/ log/ mapfile/ maplist/ pickup.sh result/
$ ls SRR794547/result
SRR794547_A.est.txt      SRR794547_DQB1.est.txt  SRR794547_DRB9.est.txt
SRR794547_A.read.txt    SRR794547_DQB1.read.txt SRR794547_DRB9.read.txt
SRR794547_B.est.txt      SRR794547_DRA.est.txt   SRR794547_E.est.txt
SRR794547_B.read.txt    SRR794547_DRA.read.txt  SRR794547_E.read.txt
SRR794547_C.est.txt      SRR794547_DRB1.est.txt  SRR794547_F.est.txt
SRR794547_C.read.txt    SRR794547_DRB1.read.txt SRR794547_F.read.txt
SRR794547_DMA.est.txt    SRR794547_DRB2.est.txt  SRR794547_G.est.txt
SRR794547_DMA.read.txt  SRR794547_DRB2.read.txt SRR794547_G.read.txt
SRR794547_DMB.est.txt    SRR794547_DRB3.est.txt  SRR794547_H.est.txt
SRR794547_DMB.read.txt  SRR794547_DRB3.read.txt SRR794547_H.read.txt
SRR794547_DOA.est.txt    SRR794547_DRB4.est.txt  SRR794547_J.est.txt
SRR794547_DOA.read.txt  SRR794547_DRB4.read.txt SRR794547_J.read.txt
SRR794547_DOB.est.txt    SRR794547_DRB5.est.txt  SRR794547_K.est.txt
SRR794547_DOB.read.txt  SRR794547_DRB5.read.txt SRR794547_K.read.txt
SRR794547_DPA1.est.txt   SRR794547_DRB6.est.txt  SRR794547_L.est.txt
SRR794547_DPA1.read.txt SRR794547_DRB6.read.txt SRR794547_L.read.txt
SRR794547_DPB1.est.txt   SRR794547_DRB7.est.txt  SRR794547_V.est.txt
SRR794547_DPB1.read.txt SRR794547_DRB7.read.txt SRR794547_V.read.txt
SRR794547_DQA1.est.txt   SRR794547_DRB8.est.txt  SRR794547_final.result.txt
SRR794547_DQA1.read.txt SRR794547_DRB8.read.txt
$
```

To be continued

Continued

```
$ cat SRR794547/result/SRR794547_final.result.txt
A   HLA-A*31:01:02   HLA-A*02:01:01
B   HLA-B*07:02:01   HLA-B*15:01:01
C   HLA-C*03:03:01   HLA-C*07:02:01
DRB1 HLA-DRB1*15:01:01   HLA-DRB1*04:04:01
DQA1 HLA-DQA1*01:02:01   HLA-DQA1*03:01:01
DQB1 HLA-DQB1*06:02:01   HLA-DQB1*03:02:01
DPA1 HLA-DPA1*01:03:01   -
DPB1 HLA-DPB1*02:01:02   HLA-DPB1*06:01
DMA  HLA-DMA*01:01:01   -
DMB  HLA-DMB*01:01:01   -
DOA  HLA-DOA*01:01:02   HLA-DOA*01:01:01
DOB  HLA-DOB*01:01:01   HLA-DOB*01:04:01
DRA  HLA-DRA*01:02:03   HLA-DRA*01:01:01
DRB2 Not typed   Not typed
DRB3 HLA-DRB3*02:02:01   -
DRB4 HLA-DRB4*01:03:01   HLA-DRB4*01:03:01
DRB5 HLA-DRB5*01:01:01   -
DRB6 HLA-DRB6*02:01   -
DRB7 HLA-DRB7*01:01:01   -
DRB8 HLA-DRB8*01:01   -
DRB9 HLA-DRB9*01:01   -
E   HLA-E*01:01:01   HLA-E*01:03:02
F   HLA-F*01:01:01   HLA-F*01:01:03
G   HLA-G*01:03:01   HLA-G*01:01:01
H   HLA-H*01:01:01   HLA-H*02:04
J   HLA-J*01:01:01   -
K   HLA-K*01:01:01   HLA-K*01:02
L   HLA-L*01:01:01   HLA-L*01:02
V   HLA-V*01:01:01   -
$
```

The file `SRR794547/result/SRR794547_final.result.txt` contains the typing results for each HLA locus if the `hlahd.sh` is executed correctly.

Hyphen means that the gene is considered as homozygous if copy number of the gene is two.