HIV.db: A package that provides HIV feature database and query APIs

Mike Jiang,Raphael Gottardo

July 20, 2012

Abstract

1 Introduction

The current HIV.db package provides simple API to access the HIV(HXB2) feature database that contains HIV genes,gene products,genomic structure elements and epitopes.

2 Load HIV feature database

`loadFeatures` function load the feature database for the selected genome and element into memory. The returned `environment` will contain the chosen reference element and its children features. The ranges of the features are given by default in amino acid coordinates relative to the reference.

```
> HIV_env<-loadFeatures(ref="env")
```

Alternately, `loadFeatures` can load features in DNA coordinates relative to the full genome.

```
> HIV_gag<-loadFeatures(ref="gag", DNA=TRUE)
```

After the database is loaded, users can explore the database by `lsCategory`, which lists the types of features availabe in the database for query.

```
> lsCategory(HIV_env)

[1] "gene"     "protein" "loop"     "RNA"      "RRE"      "binding" "strand"
[8] "helix"    "membrane"
```

3 Query HIV feature database

The package provides two query methods:`getFeature` for HIV features and `getEpitope` for HIV epitopes. Different HIV genes,proteins or viral regions can be searched by their names or categories:

1

```
> getFeature(HIV_env, name=c("gp120","gp41"))
> getFeature(HIV_env, category="loop")
> feature1<-getFeature(HIV_env, name="gp120")
```

and filter the query results further by the reading frames.

```
> getFeature(HIV_env, category=c("gene"),frame=3)

HivFeature with 1 row and 3 value columns across 1 space
      space    ranges |        name    category      frame
   <factor> <IRanges> | <character> <character> <integer>
1         1 [0, 857] |         env        gene          3
```

hxb2 coordinates can also be used to specify the start and end nucleotide positions.

```
> getFeature(HIV_gag)
> getFeature(HIV_gag,start=790,end=2000)
```

The same query can be done to the epitope database by getEpitope method execpt that it takes an extra arugment species.

```
> getEpitope(HIV_gag,start=4240,end=5096,species="mouse",name="4D6",frame=3)

Epitope with 1 row and 6 value columns across 1 space
      space       ranges |        name    category      frame        Epitope
   <factor>    <IRanges> | <character> <character> <integer>    <character>
1         1 [4353, 4392] |         4D6     Epitope         3 KCQLKGEAMHGQVD
      Species     Subtype
   <character> <character>
1       mouse           B
```

getEpitope also takes a HivFeature object as input and use the HXB2 coordinates range to get the appropriate epitopes.

```
> feature2<-getFeature(HIV_gag, name="p7")
> getEpitope(feature2, species="mouse")

Epitope with 2 rows and 6 value columns across 1 space
      space       ranges |        name    category      frame            Epitope
   <factor>    <IRanges> | <character> <character> <integer>        <character>
1         1 [2077, 2122] |         DG8     Epitope         1 RQANFLGKIWPSYKGR
2         1 [2014, 2041] |         EC6     Epitope         1       PRKKGCWKCG
      Species     Subtype
   <character> <character>
1       mouse
2       mouse           B
```

4 Query by parent/children relations

HivFeatures have the parent or children features based on the relative positions of their HXB2 coordinates. We provide two methods to query the children or parent features: `getChildren` and `getParent`.

```
> feature3<-getFeature(HIV_gag, name="gag")
> getChildren(feature3)

> feature4<-getFeature(HIV_gag, name="p24")
> getParent(feature4)

HivFeature with 1 row and 3 value columns across 1 space
      space        ranges |        name    category      frame
   <factor>    <IRanges> | <character> <character> <integer>
1         1 [790, 2292] |         gag        gene          1

> feature5<-getFeature(HIV_env, name="V1")
> getParent(feature5)

HivFeature with 1 row and 3 value columns across 1 space
      space      ranges |        name    category      frame
   <factor> <IRanges> | <character> <character> <integer>
1         1 [30, 511] |       gp120     protein          3
```

When `recursive` is set as `TRUE`,all the descendants or ancestors are returned besides the immediate children or parents.

```
> getParent(feature5,recursive=TRUE)

HivFeature with 2 rows and 3 value columns across 1 space
      space      ranges |        name    category      frame
   <factor> <IRanges> | <character> <character> <integer>
1         1 [ 0, 857] |         env        gene          3
2         1 [30, 511] |       gp120     protein          3

> getChildren(feature1,recursive=TRUE)

HivFeature with 51 rows and 3 value columns across 1 space
        space        ranges   |        name    category      frame
     <factor>    <IRanges>   | <character> <character> <integer>
1           1 [131, 156]   |          V1        loop          3
2           1 [157, 195]   |          V2        loop          3
3           1 [296, 330]   |          V3        loop          3
4           1 [385, 417]   |          V4        loop          3
5           1 [459, 469]   |          V5        loop          3
6           1 [367, 368]   |      Asp368     binding          0
```

```
7          1 [369, 370]   |      Glu370     binding       0
8          1 [426, 427]   |      Trp427     binding       0
9          1 [ 39,  41]   | beta*bar(4)     strand        3
...        ...       ... ...          ...        ...     ...
43         1 [368, 372]   |     alpha*3      helix         3
44         1 [387, 390]   |     alpha*4      helix         3
45         1 [473, 483]   |     alpha*5      helix         3
46         1 [229, 233]   |           A      loop          3
47         1 [251, 259]   |           B      loop          3
48         1 [263, 269]   |           C      loop          3
49         1 [273, 282]   |           D      loop          3
50         1 [347, 356]   |           E      loop          3
51         1 [435, 441]   |           F      loop          3
```

5   sequence of HIV feature

We also provide two methods to extract amino acid or DNA sequence of the HIV features.
Sequence can be extracted from `HivFeature` objects directly

```
> getAA(feature4)
> getDNA(feature4)
```

or by providing database environment and hxb2 coornidates range to `getAA` method:

```
> #%getAA(HIV_gag,name="pol")
>
> getDNA(HIV_gag,start=1921,end=2085)
> getDNA(feature4,start=1500,end=1615)
```

Note that both methods can take a list of `HivFeature` objects as the input and return the
their sequences respectively.

```
> getAA(getChildren(feature3))
> getDNA(getChildren(feature3))
```

6   conclusion

THe package allows users to query the built-in HIV features database for the important infor-
mation about HIV gene and gene product as well as genomic structure elements.

7   Reference

http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html