Package 'MPXV'

April 21, 2023

Title Variable Number Tandem Repeats Analysis

Version 0.1.0

Description Analysis of calling Variable Number Tandem Repeats for the genome sequence of monkeypox virus and virus tracking.

LazyData true

Imports seqinr, stringr, Biostrings

RoxygenNote 7.2.1

R topics documented:

VNTRcaller	*Calling Variable Number Tandem Repeats (VNTR) for the genome sequence of monkeypox virus (MPXV)*

Description

The function `VNTRcaller` estimates the copy numbers of VNTRs.

Usage

```
VNTRcaller(
  data,
  vntr = vntr,
  match_s = match_s,
  mismatch_s = mismatch_s,
  regionStart = regionStart,
  regionEnd = regionEnd,
  baseonly = T,
  fseq = fseq,
  fseqExtend = fseqExtend,
  VNTRoutput = F,
  tracker = F
)
```

Arguments

`data`	MPXV sequences from a file in FASTA format
`vntr`	sequence of the repetitive unit
`match_s`	matching weight
`mismatch_s`	mismatching penalty
`regionStart`	starting position of a VNTR region
`regionEnd`	ending position of a VNTR region
`baseonly`	logical. If TRUE, only the letters of the nucleotide bases (i.e. A,C,G,T) are considered. If FALSE, the degenerate codes are also considered.
`fseq`	the length of flanking sequence
`fseqExtend`	the length of extended flanking sequence
`pfseqmismatch`	tolerence proportion of number of mismatch to the length of flanking sequence
`VNTRoutput`	logical. If TRUE, the output is written to .csv files.
`tracker`	logical. If TRUE, call function [VNTRtracker]().

Value

`ID`	name of MPXV sequences
`r`	copy of tandem repeats
`match`	number of matches
`mismatch`	number of mismatches
`indel`	number of insertions and deletions (indels)
`score`	alignment score of a VNTR region for a query strain
`start_pos`	starting position of the VNTR region for a query strain
`count_n`	number of N characters
`count_deg`	number of non-ATCGN characters
`count_len`	VNTR region length
`vntr_match_norm`	
	proportion of number of matches to VNTR region length

References

Yang, H.-C. et al (2022) Monkeypox genome contains variable number tandem repeats enabling accurate virus tracking.

Pages H, Aboyoun P, Gentleman R, DebRoy S (2022) *Biostrings: Efficient manipulation of biological strings*. R package version 2.64.0.

Examples

```
## Example 1

## Read MPXV example sequences from a file in FASTA format.
MPXVseq <- read.fasta(system.file("extdata/MPXV_seq_example.fasta.gz", package = "MPXV"), as.string = T)

## VNTR
vntr <- c("T","TATGATGGA","AT","ATATACATT")
regionStart <- c(132436,150542,173240,178413)
```

```
regionEnd <- c(133216,151501,173320,179244)

## parameter settings
match_s = 2
mismatch_s = -5
baseonly = TRUE
fseq = 25
fseqExtend = 150
VNTRoutput = FALSE
tracker = FALSE

## computes the copy of the variable number tandem repeats
out <- VNTRcaller(data = MPXVseq, vntr = vntr,
                  regionStart = regionStart, regionEnd = regionEnd,
                  match_s = match_s, mismatch_s=mismatch_s,
                  baseonly = baseonly,
                  fseq = fseq, fseqExtend = fseqExtend,
                  VNTRoutput = VNTRoutput,
                  tracker = tracker)

## Example 2
MPXVseq <- read.fasta(system.file("extdata/MPXV_MT250197_Singapore.fasta.gz",
           package = "MPXV"), as.string = T)
## VNTR
vntr <- c("T","TATGATGGA","AT","ATATACATT")
regionStart <- c(132436,150542,173240,178413)
regionEnd <- c(133216,151501,173320,179244)

setwd("G:/Monkeypox/pkg_example")
## parameter settings
match_s = 2
mismatch_s = -5
baseonly = TRUE
fseq = 25
fseqExtend = 150
VNTRoutput = T
tracker = T

## computes the copy of the variable number tandem repeats
out <- VNTRcaller(data = MPXVseq, vntr = vntr,
                  regionStart = regionStart, regionEnd = regionEnd,
                  match_s = match_s, mismatch_s=mismatch_s,
                  baseonly = baseonly,
                  fseq = fseq, fseqExtend = fseqExtend,
                  VNTRoutput = VNTRoutput,
                  tracker = tracker)


## For more details about the output of VNTRcaller, please vitsit https://github.com/teresayang/MPXV_VNTR.
```

VNTRtracker   *Tracking the origin of monkeypox virus (MPXV) by using Variable Number Tandem Repeats (VNTR) markers*

Description

The function VNTRtracker calculates the MPXV strain-strain genetic distance to track the origin of a given MPXV by using VNTRs.

Usage

```
VNTRtracker(r, file_name = "Tracking_output")
```

Arguments

r   A vector of copies of VNTR of MPXV.

file_name   Filename of the output.

Details

The four VNTR markers nt133095[A25R - A26L] (T), nt150560[A47R - A49R] (TATGATGGA), nt173267[B14R - B15L] (AT), and nt179074[B18R - B19R] (ATATACATT) are proposed to track the original of MPXV strains. Based on these four VNTRs, we calculate three distances (ploymorphism information content, entropy, and length of repetitive unit (see Yang et al. (2022)) between a query MPXV strain and the strains in our database.

Value

The output contains multiple columns, including accession ID, country name, collection date, clade, lineage, and three distances mentioned above.

References

Yang, H.-C. et al (2022) Monkeypox genome contains variable number tandem repeats enabling accurate virus tracking.

Examples

```
r=c(27,7,24,16)

## The output is a csv file with the file name given by the parameter "file_name".
## In this file, each row represents a strain in our database, and column variables contain
## accession ID, country name, collection date, clade, lineage, and three distances (see Yang et al. (2022))
## between a query MPXV strain and a strains in our database.
## The rank of strains are ordered by ploymorphism information content distance (see Yang et al. (2022)) from low
VNTRtracker(r, file_name="Tracking_output")
```

Index