

Package ‘gquad’

June 7, 2017

Type Package

Title Prediction of G Quadruplexes and Other Non-B DNA Motifs

Version 2.1-1

Date 2017-06-06

Author Hannah O. Ajoge

Maintainer Hannah O. Ajoge <ohuajo@gmail.com>

Description Genomic biology is not limited to the confines of the canonical B-forming DNA duplex, but includes over ten different types of other secondary structures that are collectively termed non-B DNA structures. Of these non-B DNA structures, the G-quadruplexes are highly stable four-stranded structures that are recognized by distinct subsets of nuclear factors. This package provide functions for predicting intramolecular G quadruplexes. In addition, functions for predicting other intramolecular nonB DNA structures are included.

License Artistic-2.0

Depends R (>= 3.3.2)

Imports ape (>= 4.0), seqinr (>= 3.3-3)

LazyData TRUE

NeedsCompilation no

RoxygenNote 6.0.1

Repository CRAN

Date/Publication 2017-06-06 22:14:12 UTC

R topics documented:

aphased	2
gquad	3
gquadO	4
hdna	5
hdnaO	7
slipped	8
str	9
tfo	10
zdna	11

aphased

Predicting A-phased DNA repeat(s)

Description

This function predicts A-phased DNA repeat(s) in 'x' (DNA). DNA sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
aphased(x, xformat = "default")
```

Arguments

x	DNA sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which A-phased DNA repeat(s) will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts A-phased DNA repeat(s) in DNA sequences and provide the position, sequence and length of the predicted repeat(s), if any.

Value

A dataframe of A-phased DNA repeats' position, sequence and length. If more than one DNA sequence is provided as argument, an input ID is returned for repeat(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

Examples

```
## Predicting A-phased DNA repeat(s) from raw DNA sequences
E1 <- "TCTTGTTTTAAAACGTTTTAAAACGTTTTAAAACGTTTTAAAACGAAT"
aphased(E1)

## Predicting A-phased DNA repeat(s) from DNA sequences in fasta file
## Not run: aphased(x="Example.fasta", xformat = "fasta")

## Predicting A-phased DNA repeat(s) from DNA sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: aphased(c("BH114913", "AY611035"), xformat = "GenBank")
```

gquad	<i>Predicting G quadruplexes</i>
-------	----------------------------------

Description

This function predicts G quadruplexes in 'x' (nucleotide sequence(s)). Nucleotide sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
gquad(x, xformat = "default")
```

Arguments

x	nucleotide sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which G quadruplexes will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts G quadruplexes in nucleic (both DNA and RNA) sequences and provide the position, sequence and length of the predicted motif(s). If any motif is predicted, the degree of likeliness for the motif to be formed is computed and scored as ** (more likely) or as * (less likely).

Value

A dataframe of G quadruplexes' position, sequence, length and likeliness. If more than one nucleotide sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

See Also

gquadO

Examples

```
## Predicting G quadruplexes from raw nucleotide sequences
E1 <- c("TCTTGGGCATCTGGAGCCGGAAT", "tagtgctgggaggtagagacaggatcct")
gquad(E1)

## Predicting G quadruplexes from nucleotide sequences in fasta file
## Not run: gquad(x="Example.fasta", xformat = "fasta")

## Predicting G quadruplexes from nucleotide sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: gquad(c("BH114913", "AY611035"), xformat = "GenBank")
```

gquadO

Predicting G quadruplexes including overlaps

Description

This function predicts G quadruplexes in 'x' (nucleotide sequence(s)) like the gquad function, but includes overlaps. Nucleotide sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
gquadO(x, xformat = "default")
```

Arguments

x	nucleotide sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which G quadruplexes (including overlaps) will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts G quadruplexes in nucleic (both DNA and RNA) sequences, including overlaps and provide the position, sequence and length of the predicted motif(s). If any motif is predicted, the degree of likeliness for the motif to be formed is computed and scored as ** (more likely) or as * (less likely).

Value

A dataframe of G quadruplexes' position, sequence, length and likeliness. If more than one nucleotide sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

See Also

gquad

Examples

```
## Predicting G quadruplexes (including overlaps) from raw nucleotide sequences
E1 <- c("TCTTGGGCATCTGGAGCCGGAAT", "taggtgctgggaggtagagacaggatcct")
gquad0(E1)

## Predicting G quadruplexes (including overlaps) from nucleotide sequences in fasta file
## Not run: gquad0(x="Example.fasta", xformat = "fasta")

## Predicting G quadruplexes (including overlaps) from nucleotide sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: gquad0(c("BH114913", "AY611035"), xformat = "GenBank")
```

hdna

Predicting intramolecular triplexes (H-DNA)

Description

This function predicts H-DNA in 'x' (DNA). DNA can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
hdna(x, xformat = "default")
```

Arguments

x	DNA sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which H-DNA will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts H-DNA in DNA sequences and provide the position, sequence and length of the predicted motif(s), if any.

Value

A dataframe of H-DNA' position, sequence and length. If more than one DNA sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

See Also

hdnaO

Examples

```
## Predicting H-DNA from raw DNA sequences
E1 <- c("TCTTCCCCCTTTTTYYYYGCTYYYYTTTTTCCCCCGAAT", "tagtgctgggaggtagagacaggatcct")
hdna(E1)

## Predicting H-DNA from DNA sequences in fasta file
## Not run: hdna(x="Example.fasta", xformat = "fasta")

## Predicting H-DNA from DNA sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: hdna(c("BH114913", "AY611035"), xformat = "GenBank")
```

`hdna0`*Predicting intramolecular triplexes (H-DNA) including overlaps*

Description

This function predicts H-DNA in 'x' DNA sequence like the `hdna` function, but includes overlaps. DNA sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
hdna0(x, xformat = "default")
```

Arguments

<code>x</code>	DNA sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which H-DNA (including overlaps) will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
<code>xformat</code>	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts H-DNA in DNA sequences, including overlaps and provide the position, sequence and length of the predicted motif(s), if any.

Value

A dataframe of H-DNA' position, sequence and length. If more than one DNA sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

See Also

`hdna`

Examples

```
## Predicting H-DNA (including overlaps) from raw DNA sequences
E1 <- c("TCTTCCCCCTTTTYYYYGCTYYYYTTTTCCCCCGAAT", "taggtgctgggagtagagacaggatcct")
hdna0(E1)

## Predicting H-DNA (including overlaps) from DNA sequences in fasta file
## Not run: hdna0(x="Example.fasta", xformat = "fasta")

## Predicting H-DNA (including overlaps) from DNA sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: hdna0(c("BH114913", "AY611035"), xformat = "GenBank")
```

slipped	<i>Predicting slipped motif(s)</i>
---------	------------------------------------

Description

This function predicts slipped motif(s) in 'x' in DNA. DNA sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
slipped(x, xformat = "default")
```

Arguments

x	DNA sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which slipped motif(s) will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts slipped motif(s) in DNA sequences and provide the position, sequence and length of the predicted motif(s). If any motif is predicted, the degree of likeliness for the motif to be formed is computed and scored as ** (more likely) or as * (less likely).

Value

A dataframe of slipped motif(s) position, sequence, length and likeliness. If more than one DNA sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

Examples

```
## Predicting slipped motif(s) from raw DNA sequences
E1 <- c("TCTTACTGTGACTGTGGAAT", "tagtgctgggaggtagagacaggatcct")
slipped(E1)

## Predicting slipped motif(s) from DNA sequences in fasta file
## Not run: slipped(x="Example.fasta", xformat = "fasta")

## Predicting slipped motif(s) from DNA sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: slipped(c("BH114913", "AY611035"), xformat = "GenBank")
```

str

Predicting short tandem repeats

Description

This function predicts short tandem repeats in 'x' in nucleotides. Nucleotide sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
str(x, xformat = "default")
```

Arguments

x	Nucleotide sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which short tandem repeats will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts short tandem repeats in nucleotide sequences and provide the position, sequence and length of the predicted repeats, if any.

Value

A dataframe of short tandem repeats' position, sequence and length. If more than one DNA sequence is provided as argument, an input ID is returned for repeats predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

Examples

```
## Predicting short tandem repeats from raw nucleotide sequences
E1 <- c("TCTACACACACACACACACGAAT", "tagggugugugugugugugugutcct")
str(E1)

## Predicting short tandem repeats from nucleotide sequences in fasta file
## Not run: str(x="Example.fasta", xformat = "fasta")

## Predicting short tandem repeats from nucleotide sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: str(c("BH114913", "AY611035"), xformat = "GenBank")
```

tfo

Predicting triplex forming oligonucleotide(s)

Description

This function predicts triplex forming oligonucleotide(s) in 'x' in DNA. DNA sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
tfo(x, xformat = "default")
```

Arguments

x	DNA sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which triplex forming oligonucleotide(s) will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts triplex forming oligonucleotide(s) in DNA sequences and provide the position, sequence and length of the predicted motif(s), if any.

Value

A dataframe of triplex forming oligonucleotide(s) position, sequence and length. If more than one DNA sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

Examples

```
## Predicting triplex forming oligonucleotide(s) from raw DNA sequences
E1 <- c("TCTTGGGAGGGAGAGAGAGAAAGAGATCTGGAGGCCGGAAT", "taggtgctggaggtagagacaggatcct")
tfo(E1)

## Predicting triplex forming oligonucleotide(s) from DNA sequences in fasta file
## Not run: tfo(x="Example.fasta", xformat = "fasta")

## Predicting triplex forming oligonucleotide(s) from DNA sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: tfo(c("BH114913", "AY611035"), xformat = "GenBank")
```

zdna

Predicting Z-DNA motif(s)

Description

This function predicts Z-DNA motif(s) in 'x' in DNA. DNA sequence can be provided in raw or fasta format or as GenBank accession number(s). Internet is needed to connect to GenBank database, if accession number(s) is given as argument.

Usage

```
zdna(x, xformat = "default")
```

Arguments

x	DNA sequence(s) in raw format or a fasta file or a GenBank accession number(s); from which Z-DNA motif(s) will be predicted. If the fasta file name does not contain an absolute path, the file name is relative to the current working directory.
xformat	a character string specifying the format of x : default (raw), fasta, GenBank (GenBank accession number(s)).

Details

This function predicts Z-DNA motif(s) in DNA sequences and provide the position, sequence and length of the predicted motif(s). If any motif is predicted, the degree of likeliness for the motif to be formed is computed and scored as ** (more likely) or as * (less likely).

Value

A dataframe of Z-DNA motif(s) position, sequence, length and likeliness. If more than one DNA sequence is provided as argument, an input ID is returned for motif(s) predicted from each input sequence.

Author(s)

Hannah O. Ajoge

References

paper under review

Examples

```
## Predicting Z-DNA motif(s) from raw DNA sequences
E1 <- c("TCTTGCGCGCGCGCGCGCGCAAT", "taggtgctgggaggtagagacaggatcct")
zdna(E1)

## Predicting Z-DNA motif(s) from DNA sequences in fasta file
## Not run: zdna(x="Example.fasta", xformat = "fasta")

## Predicting Z-DNA motif(s) from DNA sequences,
## using GenBank accession numbers.
## Internet connectivity is needed for this to work.
## Not run: zdna(c("BH114913", "AY611035"), xformat = "GenBank")
```

Index

aphased, [2](#)

gquad, [3](#)
gquad0, [4](#)

hdna, [5](#)
hdna0, [7](#)

slipped, [8](#)
str, [9](#)

tfo, [10](#)

zdna, [11](#)