Package: gquad (via r-universe)

October 21, 2024

Type Package

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

Version 2.1-2

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 (>= 4.2.0)

Imports ape (>= 5.6-2), seqinr (>= 4.2-23)

Encoding UTF-8

RoxygenNote 7.2.2

Suggests knitr, rmarkdown

VignetteBuilder knitr

NeedsCompilation no

Date/Publication 2022-11-29 08:40:02 UTC

Repository https://ohuajo.r-universe.dev

RemoteUrl https://github.com/cran/gquad

RemoteRef HEAD

RemoteSha 6466c382c945ac6cb15e67fff50263eed4e97458

Contents

---

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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

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   *Predicting G quadruplexes*

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 nu-
cleotide 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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vi-
gnettes

See Also

gquadO

Examples

```
## Predicting G quadruplexes from raw nucleotide sequences
E1 <- c("TCTTGGGCATCTGGAGGCCGGAAT", "taggtgctgggaggtagagacaggatatcct")
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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

See Also

gquad

Examples

```
## Predicting G quadruplexes (including overlaps) from raw nucleotide sequences
E1 <- c("TCTTGGGCATCTGGAGGCCGGAAT", "taggtgctgggaggtagagacaggatatcct")
gquadO(E1)

## Predicting G quadruplexes (including overlaps) from nucleotide sequences in fasta file
## Not run: gquadO(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: gquadO(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 num-
                ber(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 direc-
                tory.

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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vi-
gnettes

See Also

hdnaO

Examples

```
## Predicting H-DNA from raw DNA sequences
E1 <- c("TCTTCCCCCCTTTTTYYYYYGCTYYYYYTTTTTCCCCCCGAAT", "taggtgctgggaggtagagacaggatatcct")
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")
```

---

hdnaO   *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

```
hdnaO(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 (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 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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

See Also

hdna

Examples

```
## Predicting H-DNA (including overlaps) from raw DNA sequences
E1 <- c("TCTTCCCCCCTTTTTYYYYYGCTYYYYYTTTTTCCCCCCGAAT", "taggtgctgggaggtagagacaggatatcct")
hdnaO(E1)

## Predicting H-DNA (including overlaps) from DNA sequences in fasta file
## Not run: hdnaO(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: hdnaO(c("BH114913", "AY611035"), xformat = "GenBank")
```

---

 slipped   *Predicting slipped motif(s)*

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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

Examples

```
 ## Predicting slipped motif(s) from raw DNA sequences
E1 <- c("TCTTACTGTGACTGTGGAAT", "taggtgctgggaggtagagacaggatatcct")
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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

Examples

```
 ## Predicting short tandem repeats from raw nucleotide sequences
E1 <- c("TCTACACACACACACACACACGAAT", "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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

Examples

```
 ## Predicting triplex forming oligonucleotide(s) from raw DNA sequences
E1 <- c("TCTTGGGAGGGAGAGAGAGAAAGAGATCTGGAGGCCGGAAT", "taggtgctgggaggtagagacaggatatcct")
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 num-
            ber(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 on gquad and the web application (Non-B DNA Predictor) is under review, see draft in vignettes

Examples

```
 ## Predicting Z-DNA motif(s) from raw DNA sequences
E1 <- c("TCTTGCGCGCGCGCGCGCGCGCGCGCAAT", "taggtgctgggaggtagagacaggatatcct")
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