Package 'VDJgermlines'

October 12, 2022

Type Package

Title Variable, Diversity and Joining Sequences from Various Species

Version 0.1

Date 2018-12-10

Author Alexander Yermanos

Maintainer Alexander Yermanos <ayermanos@gmail.com>

Depends R(>= 3.1.0), ape, stringdist

Description Contains variable, diversity, and joining sequences and accompanying functions that enable both the extraction of and comparison between immune V-D-J genomic segments from a variety of species. Sources include IMGT from MP Lefranc (2009) <doi:10.1093/nar/gkn838> and Vgenerepertoire from publication DN Olivieri (2014) <doi:10.1007/s00251-014-0784-3>.

License GPL-2

LazyData true

Encoding UTF-8

RoxygenNote 5.0.1

Suggests knitr, rmarkdown

NeedsCompilation no

Repository CRAN

Date/Publication 2018-12-18 23:30:32 UTC

R topics documented:

---

extractSequencesR   *Extracts the germline genes of interest and stores them as a dataframe*

Description

Extracts the germline genes of interest and stores them as a dataframe

Usage

```
extractSequencesR(species, chain, source, write.fasta, fasta.name,
  directory.string)
```

Arguments

species   Specifies desired species. Current species options can be found by running
                   unique(VDJgermlines$species) after loading in data(VDJgermlines). "all_species"
                   will select all available species. Species include ("alpaca","bovine","crab_eating_macaque","dog",
                   "human","mouse","pig","platypus","rabbit",raintrout","rat","rhesus monkey","salmon",
                   "sheep","zebrafish","atlanticcod","catfish","teleostei","goat","dolphin","mamonkey",
                   "camel","runny")

chain   Specifies the desired chain. "all_chains" will select all chains. The various
                   chains can be listed by running unique(VDJgermlines$chain).Antibody loci start
                   with IG, TCR starts with TR.NExt comes either heavy, light, alpha, beta etc.
                   Examples include IGHD IGHJ IGHV IGIJ IGIV IGKJ IGKV IGLJ IGLV TRAJ
                   TRAV TRBD TRBJ TRBV TRDD TRDJ TRDV TRGJ TRGV

source   Specifies from which source the desired chains should be taken from. Current
                   sources can be found by running unique(VDJgermlines$source) after loading in
                   data(VDJgermlines). IMGT and vgenerepertoire are currently present.

write.fasta   A boolean (TRUE / FALSE) value that species if the output sequences should
                   be also written as a fasta file.

fasta.name   A string specifying the name of the fasta file. If this is left blank this will be
                   defaulted to "VDJgermlines.fasta". Manually including extension is not needed.

directory.string
                   A string stating the directory that the fasta file should be saved to. If left blank
                   the fasta file will be stored in the current working directory.

Value

Returns a dataframe containing the species and chains of interest from the desired source. The
columns of the dataframe include sequence, names, species, chain, accession, official species name,
source, and strain. Furthermore, if write.fasta is set to TRUE then the output will additionally
include a fasta file.

Examples

```
extractSequencesR(species = "dog",chain = "IGHD",source = "IMGT",write.fasta = FALSE)
```

| VDJgermlines | *VDJgermlines* |

Description

The germline information for a variety of species is contained in the VDJgermlines dataframe.

Usage

```
VDJgermlines
```

Format

A data frame with 7285 rows and 8 variables:

sequence The sequence directly from either IMGT or Vgenerepertoire

names The corresponding name of the sequence including V gene number, species

species The casual species name - e.g. dog, mouse, rat

chain The corresponding chain names, including IGHV, IGHL, TRAV, etc

accession The accession number of the given gene

species_name_official The more scientific species name e.g. mus musculus

source The database that annotated the sequence

strain The accompanying strain if available (e.g. B6 vs BALB/c)

Source

IMGT and Vgenerepertoire IMGT (the international ImMunoGeneTics information system (founder and director: Marie-Paule Lefranc, Montpellier, France)." : Lefranc, M.-P. et al., Nucleic Acids Research, 27, 209-212 (1999) Cover; Ruiz, M. et al., Nucleic Acids Research, 28, 219-221 (2000); Lefranc, M.-P., Nucleic Acids Research, 29, 207-209 (2001); Lefranc, M.-P., Nucleic Acids Res., 31, 307-310 (2003); Lefranc, M.-P. et al., In Silico Biol., 5, 0006 (2004) [Epub], 5:45-60 (2005); Lefranc, M.-P. et al., Nucleic Acids Res., 33, D593-D597 (2005) Full text; Lefranc, M.-P. et al., Nucleic Acids Research 2009 37(Database issue): D1006-D1012; doi:10.1093/nar/gkn838 Full text. Nucleic Acids Res. 2015 Jan 28;43(Database issue):D413-422. doi: 10.1093/nar/gku1056. Epub 2014 Nov 5. Full text) and Vgenerepertoire (Immunogenetics. 2014 Aug;66(7-8):479-92. doi: 10.1007/s00251-014-0784-3. Epub 2014 Jun 4. and Vgenerepertoire (Genomic V exons from whole genome shotgun data in reptiles. Olivieri DNx, von Haeften B, S<U+00E1>nchez-Espinel C, Faro J, Gamb<U+00F3>n-Deza F.).

VDJheatmap	*Calculates the distance between VDJ germlines of interest and produces a matrix that can be used for further analysis or viewed using Heatmap/pheatmap/ComplexHeatmap.*

Description

Calculates the distance between VDJ germlines of interest and produces a matrix that can be used for further analysis or viewed using Heatmap/pheatmap/ComplexHeatmap.

Usage

```
VDJheatmap(species, chain, source, inference.method)
```

Arguments

species	Specifies desired species. Current species options can be found by running unique(VDJgermlines$species) after loading in data(VDJgermlines). "all_species" will select all available species. Species include ("alpaca","bovine","crab_eating_macaque","dog", "human","mouse","pig","platypus","rabbit",raintrout","rat","rhesus monkey","salmon", "sheep","zebrafish","atlanticcod","catfish","teleostei","goat","dolphin","mamonkey", "camel","runny")
chain	Specifies the desired chain. "all_chains" will select all chains. The various chains can be listed by running unique(VDJgermlines$chain).Antibody loci start with IG, TCR starts with TR.NExt comes either heavy, light, alpha, beta etc. Examples include IGHD IGHJ IGHV IGIJ IGIV IGKJ IGKV IGLJ IGLV TRAJ TRAV TRBD TRBJ TRBV TRDD TRDJ TRDV TRGJ TRGV
source	Specifies from which source the desired chains should be taken from. Current sources can be found by running unique(VDJgermlines$source) after loading in data(VDJgermlines). IMGT and vgenerepertoire are currently present.
inference.method	
	Specifies the distance metric to be used to calculate distance. Currently includes parameters from stringdist::stringdistmatrix(), such as "lv","jv","hamming","qgram","cosine","Jaccard","! or "soundex".

Value

Returns a matrix where the entries correspond to the VDJ germlines used the neighbor joining algorithm. Can be viewed by heatmap() in base r, or pheatmap::pheatmap(VDJheatmap).

Examples

```
VDJheatmap(species = "dog",chain = "IGHD",source = "IMGT")
```

VDJphylo	*Infers basic phylogenetic trees of VDJ germlines of interest with basic distance methods.*

Description

Infers basic phylogenetic trees of VDJ germlines of interest with basic distance methods.

Usage

```
VDJphylo(species, chain, source, write.fasta, fasta.name, directory.string,
  inference.method)
```

Arguments

species	Specifies desired species. Current species options can be found by running unique(VDJgermlines$species) after loading in data(VDJgermlines). "all_species" will select all available species. Species include ("alpaca","bovine","crab_eating_macaque","dog", "human","mouse","pig","platypus","rabbit",raintrout","rat","rhesus monkey","salmon", "sheep","zebrafish","atlanticcod","catfish","teleostei","goat","dolphin","mamonkey", "camel","runny")
chain	Specifies the desired chain. "all_chains" will select all chains. The various chains can be listed by running unique(VDJgermlines$chain).Antibody loci start with IG, TCR starts with TR.NExt comes either heavy, light, alpha, beta etc. Examples include IGHD IGHJ IGHV IGIJ IGIV IGKJ IGKV IGLJ IGLV TRAJ TRAV TRBD TRBJ TRBV TRDD TRDJ TRDV TRGJ TRGV
source	Specifies from which source the desired chains should be taken from. Current sources can be found by running unique(VDJgermlines$source) after loading in data(VDJgermlines). IMGT and vgenerepertoire are currently present.
write.fasta	A boolean (TRUE / FALSE) value that species if the output sequences should be also written as a fasta file.
fasta.name	A string specifying the name of the fasta file. If this is left blank this will be defaulted to "VDJgermlines.fasta". Manually including extension is not needed.
directory.string	
	A string stating the directory that the fasta file should be saved to. If left blank the fasta file will be stored in the current working directory.
inference.method	
	Specifies the method to be used to infer the phylogenetic tree. Currently includes "lv" for Levenshtein distance.

Value

Returns an unrooted tree inferred using the neighbor joining algorithm. Can be viewed by plot(output_tree) after loading ape package.

Examples

```
VDJphylo(species = "dog",chain = "IGHD",source = "IMGT",write.fasta = FALSE)
```

Index