Package 'NuPoP'

August 29, 2010

Type Package

Title An R package for nucleosome positioning prediction

Version 1.0

Date 2010-06-24

Author Ji-Ping Wang <jzwang@northwestern.edu>; Liqun Xi <lxi700@northwestern.edu>

Maintainer Ji-Ping Wang<jzwang@northwestern.edu>

Description NuPoP is an R package for Nucleosome Positioning Prediction.This package is built upon a duration hidden Markov model proposed in Xi et al, 2010; Wang et al, 2008. The core of the package was written in Fotran. In addition to the R package, a stand-alone Fortran software tool is also available at http://nucleosome.stats.northwestern.edu.

License GPL-2

R topics documented:

`NuPoP-package` *An R package for nucleosome positioning prediction*

Description

`NuPoP` is an R package for `Nu`cleosome `Po`sitioning `P`rediction. This package is built upon a duration hidden Markov model proposed in Xi et al 2010 and Wang et al 2008. The core of the package was written in Fortran. Three functions including `predNuPoP`, `readNuPoP`, and `plotNuPoP` are provided for nucleosome positioning prediction, prediction results readin, and prediction results visualization respectively. The input DNA sequence can be of any length.

Details

Package:	NuPoP
Type:	Package
Version:	1.0
Date:	2010-06-24
License:	GPL-2

`predNuPoP`: R function invoking Fortran codes to predict nucleosome positioning, nucleosome occupancy and binding affinity.

`readNuPoP`: R function to read in the prediction results by `predNuPoP`.

`plotNuPoP`: R function to visualize predictions.

Author(s)

Ji-Ping Wang, Liqun Xi

Maintainer: Ji-Ping Wang<jzwang@northwestern.edu>

References

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J. and Wang, J.-P. (2010), Predicting nucleosome positioning using a duration Hidden Markov Model, BMC Bioinformatics , doi:10.1186/1471-2105-11-346

Wang, J.-P., Fondufe-Mittendorf, Y., Xi, L., Tsai, G., Segal, E. and Widom, J.(2008), Preferentially quantized linker DNA lengths in Saccharomyces cerevisiae, PLoS Computational Biology, 4(9) e1000175

Examples

```
library(NuPoP)
predNuPoP(system.file("extdata", "test.seq", package="NuPoP"),species=7,model=4)
#temp=readNuPoP("test.seq_Prediction4.txt",startPos=1,endPos=5000)
#plotNuPoP(temp)
```

`plotNuPoP` *R function for plotting the predicted nucleosome positioning map and nucleosome occupancy map*

Description

This function produces two plots from a specified region based on the prediction results from function `predNuPoP`. The first plot is the nucleosome occupancy (grey color). In the second plot, in addition to the occupancy, Viterbi prediction (red rectangle) and the posterior probability for a position to be the start of a nucleosome (blue color) are superimposed.

Usage

```
plotNuPoP(predNuPoPResults)
```

Arguments

`predNuPoPResults`
: NuPoP prediction results from predNuPoP function. It must be a data frame read in by readNuPoP function.

Value

`plotNuPoP` outputs two plots: the nucleosome occupancy score map and Viterbi optimal nucleosome positioning map together with posterior probability for a position to be the start of a nucleosome.

Examples

```
library(NuPoP)
#predNuPoP(system.file("extdata", "test.seq", package="NuPoP"),species=7,model=4)
#temp=readNuPoP("test.seq_Prediction4.txt",startPos=1,endPos=5000)
#plotNuPoP(temp)
```

---

predNuPoP *R function for nucleosome positioning prediction, occupancy score and nucleosome binding affinity score calculation*

Description

This function invokes Fortran codes to compute the Viterbi prediction of nucleosome positioning, nucleosome occupancy score and nucleosome binding affinity score . A pre-trained linker DNA length distribution for the current species is used in a duration Hidden Markov model.

Usage

```
predNuPoP(file,species=7,model=4)
```

Arguments

`file`
: a string for the path and name of a DNA sequence file in FASTA format. This sequence file can be located in any directory. It must contain only one sequence of any length. By FASTA format, we require each line to be of the same length (the last line can be shorter; the first line should be '>sequenceName'). The length of each line should be not longer than 10 million bp.

`species`
: an integer from 0 to 11 as the label for a species indexed as follows: 1 = Human; 2 = Mouse; 3 = Rat; 4 = Zebrafish; 5 = D. melanogaster; 6 = C. elegans; 7 = S. cerevisiae; 8 = C. albicans; 9 = S. pombe; 10 = A. thaliana; 11 = Maize; 0 = Other. The default is 7 = S. cerevisiae . If `species=0` is specified, NuPoP will identify a species from 1-11 that has most similar base composition to the input sequence, and then use the models from the selected species for prediction.

`model`
: an integer = 4 or 1. NuPoP has two models integrated. One is the first order Markov chain for both nucleosome and linker DNA states. The other is 4th order (default). The latter distinguishes nucleosome/linker in up to 5-mer usage, and thus is slightly more effective in prediction, but runs slower. The time used by 4th order model is about 2.5 times of the 1st order model.

Value

predNuPoP outputs the prediction results into the current working directory. The output file is named after the input file with an added extension _Prediction1.txt or _Prediction4.txt, where 1 or 4 stands for the order of Markov chain models specified. The output file has five columns, Position, P-start, Occup, N/L, Affinity:

Position	position in the input DNA sequence
P-start	probability that the current position is the start of a nucleosome
Occup	nucleosome occupancy score
N/L	nucleosome (1) or linker (0) for each position based on Viterbi prediction
Affinity	nucleosome binding affinity score

Examples

```
library(NuPoP)
predNuPoP(system.file("extdata", "test.seq", package="NuPoP"),species=7,model=4)
```

---

readNuPoP *R function for plotting the predicted nucleosome positioning map and nucleosome occupancy map*

Description

This function reads in the prediction results generated by predNuPoP for specified region.

Usage

```
readNuPoP(file,startPos,endPos)
```

Arguments

file	the prediction output file name from predNuPoP function.
startPos	the start position in the DNA sequence for prediction results plotting.
endPos	the end position in the DNA sequence for prediction results plotting.

Examples

```
#library(NuPoP)
#predNuPoP(system.file("extdata", "test.seq", package="NuPoP"),species=7,model=4)
#temp=readNuPoP(test.seq_Prediction4.txt,startPos=1,endPos=5000)
```

Index