CNVSelectR Example

Peter Chi

5/7/2021

Preliminaries

This document demonstrates how to use the `CNVSelectR` package to take input files and run the method on them. First, we load the package:

```
library(CNVSelectR)
```

```
## Loading required package: Matrix
```

```
## Loading required package: seqinr
```

```
## Warning: package 'seqinr' was built under R version 4.0.5
```

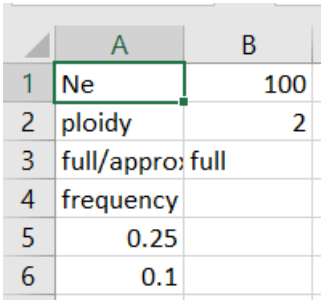
```
## Loading required package: readr
```

```
## Loading required package: knitr
```

Requirements

Two input files are required:

- csv file containing two columns, as shown below:



	A	B
1	Ne	100
2	ploidy	2
3	full/appro:	full
4	frequency	
5	0.25	
6	0.1	

Figure 1: example csv file

- txt file containing aligned sequences in FASTA format

It is assumed that these files are in your current working directory.

Running the method

Now, to generate the null model and obtain confidence intervals and p-values for each duplicate pair, we run:

```
test_out <- CNVSelect_test("cnv_sample_file_1.csv", "cnv_sample_file_2.txt")
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   X1 = col_character(),
##   X2 = col_character()
## )
```

This output looks as follows:

```
test_out
```

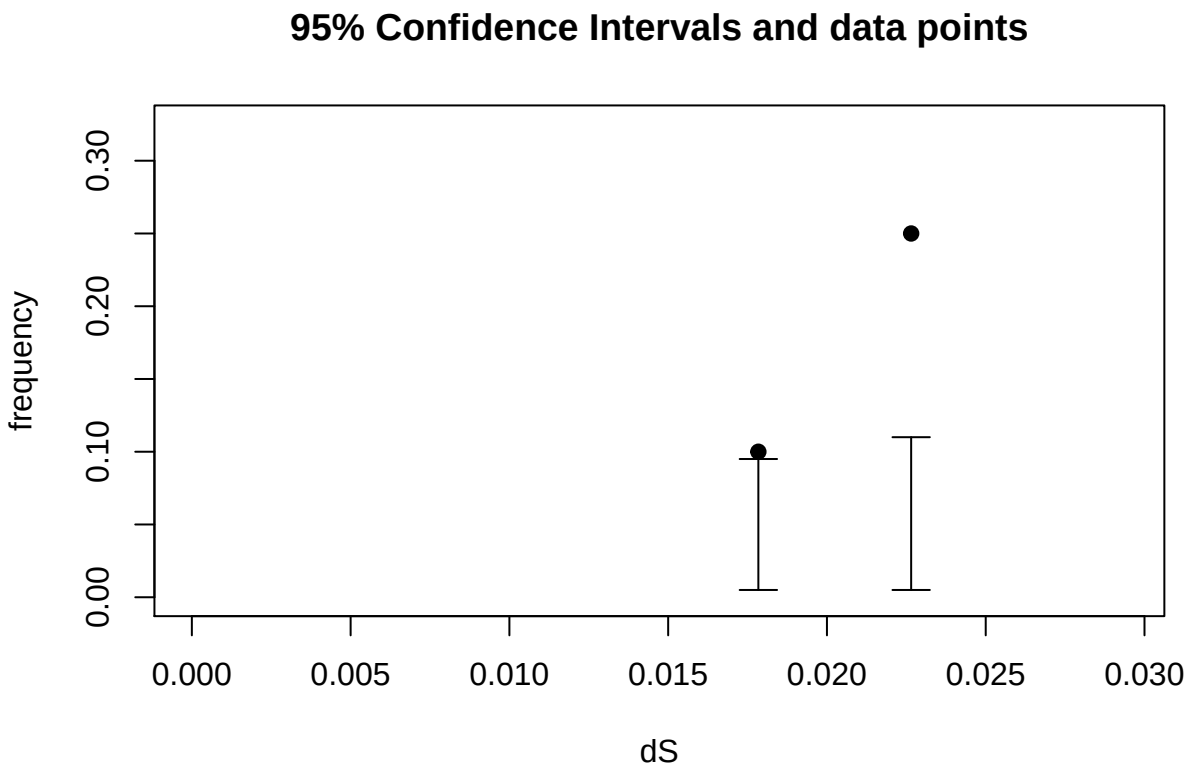
```
## $freqs
## [1] 0.25 0.10
##
## $dS
##        wt1/dup1   wt2/dup2
## [1,] 0.02265236 0.01784061
##
## $CIlower
## [1] 0.005 0.005
##
## $CIupper
## [1] 0.110 0.095
##
## $p_val
## [1] 0.0002111328 0.0361783105
```

We can create a summary table and plot as follows:

```
CNVSelect_summary(test_out)
```

	dS	frequency	95% CI	p-value
wt1/dup1	0.0227	0.25	(0.005, 0.11)	0.000211
wt2/dup2	0.0178	0.1	(0.005, 0.095)	0.0362

```
CNVSelect_plot(test_out)
```


95% Confidence Intervals and data points
frequency
0.00
0.10
0.20
0.30
0.000
0.005
0.010
0.015
0.020
0.025
0.030
dS