dnaplotr package example (version 0.1)

Scott Sherrill-Mix

Abstract

This is a collection of examples of usage for the **dnaplotr** package.

Keywords: visualization, alignment, display, genome, DNA, sequence, multiple sequence alignment, reference sequence.

---

1. General description

dnaplotr allows rapid visual assessment of many DNA, RNA or amino acid sequences by plotting each sequence as a row of colors with each color representing a base/amino acid. A simple example of using the **dnaplotr** package in R (Figure 1):

```
> set.seed(1234)
> fakeSeqs<-createFakeDNA(5000,500)
> refSeq<-fakeSeqs[1]
> fakeSeqs<-fakeSeqs[-1]
> species<-sprintf('Species %s',sub(' [0-9]+$','',names(fakeSeqs)))
> par(mar=c(3.5,4.4,.5,7),mgp=c(2.5,1,0))
> plotDNA(fakeSeqs,groups=species)
```

The package is also useful to visualize high-throughput, short read data in a given region (Figure 2):

```
> set.seed(1234)
> seqLength<-1000
> fakeSeqs<-createFakeDNA(5000,seqLength,pGap=0)
> refSeq<-fakeSeqs[1]
> fakeSeqs<-fakeSeqs[-1]
> potentialStarts<-1:(seqLength-99)
> #leave a gap in sequences
> potentialStarts<-potentialStarts[potentialStarts>550|potentialStarts<400]
> startCoords<-sort(sample(potentialStarts,5000,TRUE))
> endCoords<-startCoords+99
> dummy<-paste(rep('-',seqLength),collapse='')
> substring(fakeSeqs,1,startCoords)<-substring(dummy,1,startCoords)
> substring(fakeSeqs,endCoords+1,seqLength)<-substring(dummy,endCoords+1,seqLength)
> fakeSeqs<-replaceOuterGaps(fakeSeqs)
> par(mar=c(3.5,4.4,.5,1),mgp=c(2.5,1,0))
> plotDNA(fakeSeqs)
```

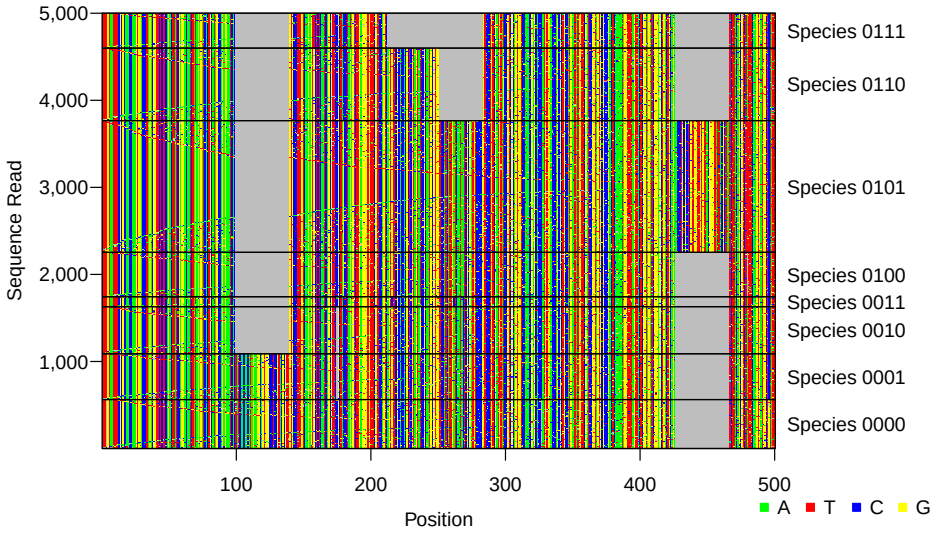

Figure 1: An example of a comparison of many sequences aligned against a reference

Amino acids can also be plotted although the larger number of amino acids makes comparisons somewhat more difficult (Figure 3). This is somewhat ameliorated by the use of a color scheme based on Jmol where amino acids with similar characteristics are colored similarly.

```
> fakeAAs<-createFakeAA()
> par(mar=c(3.5,4.4,.5,7),mgp=c(2.5,1,0))
> plotAA(fakeAAs)
```

Affiliation:

Github: http://github.com/sherrillmix/dnaplotr

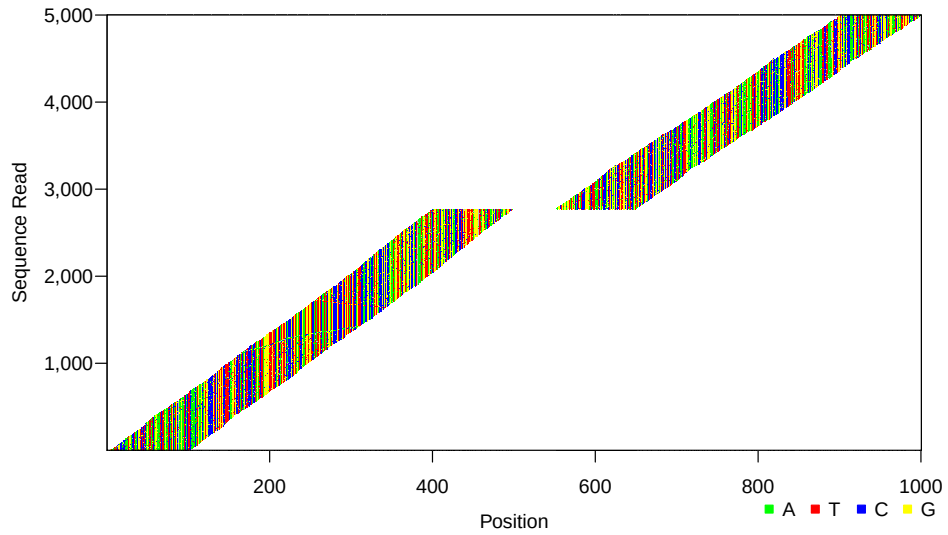

Figure 2: An example plot of short reads aligned against a reference sequence. The gap in coverage might indicate an indel, mapping problem or incorrect reference.

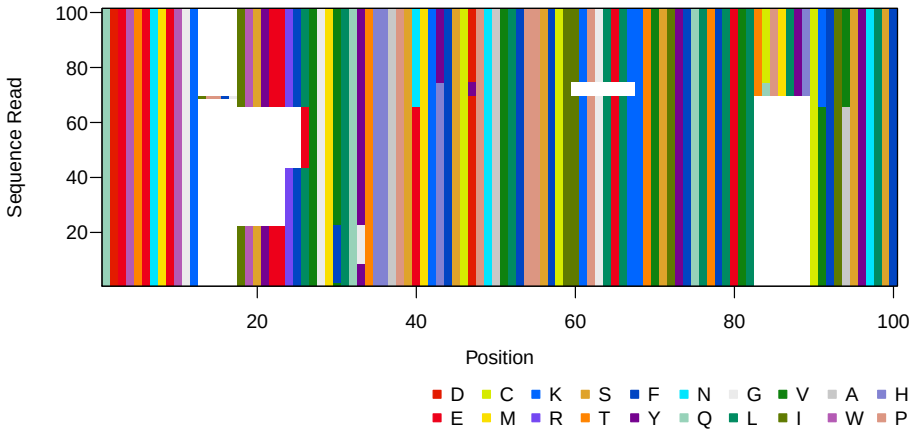

Figure 3: An example plot of amino acids