

Course Code: BT 4019

Course Title: Statistical Methods in Bioinformatics

## **Home Work 01 - Report**

Name : Nimna Alupotha Gamage (NIMNA A. G. T.)

Index No.: s14682

Reg. No. : 2019s17241

Degree : Bioinformatics

Year : 4<sup>th</sup> year

HW1.R

User-Nimna

2023-07-03

*#BT 4019 - Statistical Methods in Bioinformatics*

*#Home Work 01*

*#Date: 02.07.2023*

*#Question-Write an R function to return the complement of a given DNA sequence.*

*#Answers*

*#1-An unoptimized code with if clause inside a one for loop*

```
complement = function(sequence1){  
  for(base in sequence1)  
    if(base == "A"){  
      print("T")  
    }else if(base == "T"){  
      print("A")  
    }else if(base == "G"){  
      print("C")  
    }else if(base == "C"){  
      print("G")  
    }else{  
      print("N")  
    }  
}
```

*#calling the function 'complement'*

```
complement(c("A", "C", "G", "T", "G", "A"))
```

```
## [1] "T"
```

```
## [1] "G"
```

```
## [1] "C"
```

```
## [1] "A"
```

```
## [1] "C"
```

```
## [1] "T"
```

*#erros-does not give the output in a single line*

*# -does not give the complement, when the sequence is given in lower case letters*

*#2-An somewhat optimized code with 'gsub' function*

```
get_complement_dna1 <- function(sequence2) {  
  comp_sequence <- gsub("A", "t", sequence2)  
  comp_sequence <- gsub("T", "a", comp_sequence)  
  comp_sequence <- gsub("C", "g", comp_sequence)  
  comp_sequence <- gsub("G", "c", comp_sequence)  
  comp_seq <- toupper(comp_sequence)  
  return(comp_seq)  
}
```

*#calling the function 'get\_complement\_dna1'*

```
get_complement_dna1(c("A","C","G","T","G","A"))
```

```
## [1] "T" "G" "C" "A" "C" "T"
```

```
get_complement_dna1(c("a","C","g","T","G","p"))
```

```
## [1] "A" "G" "G" "A" "C" "P"
```

*#errors--does not give the complement, when the sequence is given in lower case letters*

*#The best optimized code*

*#3- An optimized code without using a for loop*

```
get_complement_dna = function(dna_sequence){  
  dna_seq= toupper(strsplit(dna_sequence,""))  
  comp_dna= ifelse(dna_seq== "A", "T",  
                  ifelse(dna_seq == "T", "A",  
                        ifelse(dna_seq == "G", "C",  
                              ifelse(dna_seq == "C", "G", "N"))))  
  
  complement_dna=paste ((comp_dna), collapse = "")  
  cat("The complement of", dna_seq, "DNA sequence is", complement_dna)  
}
```

*#calling the function 'get\_complement\_dna'*

```
get_complement_dna(c("A","C","G","T","G","A"))
```

```
## The complement of A C G T G A DNA sequence is TGCACT
```

```
get_complement_dna(c("A","T","S"))
```

```
## The complement of A T S DNA sequence is TAN
```

```
get_complement_dna(c("a","C","g","T","G","p"))
```

```
## The complement of A C G T G P DNA sequence is TGCACN
```