

Homework #1: Human Mitochondrion

Anže Pečar (63060257)

October 23, 2012

1 Introduction

In the first homework we analyse the human mitochondrial DNA sequence.

2 Data

Data was obtained from GenBank and is given in FASTA format. In FASTA nucleotides are represented using single letter codes (A, T, C and G). FASTA format defines many other letters of which only the letter N was present in our data. N can represent aNy nucleotide, but because there was only one such occurrence, we simply ignore it.

3 Methods

4 Results

4.1 Probabilities for k-mers

```
C - 31.29%, A - 30.97%, T - 24.69%, G - 13.05%,  
CC - 10.71%, AA - 9.72%, CA - 9.26%, AC - 9.03%, CT - 8.70%, TA - 8.30%,  
AT - 7.42%, TC - 7.25%, TT - 6.05%, AG - 4.81%, GC - 4.30%, GA - 3.70%,  
TG - 3.09%, CG - 2.62%, GT - 2.52%, GG - 2.52%,
```

Honor Code

My answers to homework are my own work. I did not make solutions or code available to anyone else. I did not engage in any other activities that will dishonestly improve my results or dishonestly improve/hurt the results of others.