

Finding repetition in genomes

Report Name	Finding repetition in genomes
Author (User Id)	Raesah Khan (rak12)
Supervisor (User Id)	Amanda Clare (afc)

Module	CS39440
Degree Scheme	G400 (Computer Science)

Date	February 4, 2023
Revision	1.0
Status	Draft

1 Project description

The goal of this project is to find microsatellites in genomes. Microsatellites are repetitive sections of DNA that repeat throughout the genome. This project should take into account microsatellites that have slight deviations from the microsatellites ie. ATG and AGT may be counted as the same microsatellite but the score may have a penalty to reflect the deviation. DNA is stored as strings, using string manipulation bioinformaticians can analyse the DNA and are then able to make conclusions about the DNA collected. This project aims to use appropriate data structures and algorithms to find these microsatellites and then to output the results in an appropriate format, this project may showcase the data using visual techniques. For the use of this project, a user should be able to input a FASTA file and give a certain k-mer size and penalty score for deviations that may be found. FASTA file

2 Proposed tasks

Research python and c for project

- Research micro satellites

- Research legacy software that find micro satellites

- Research suitable visual formats the data can be presented as

- Create code to read FASTA file

- Create code that inputs data in a GFF file format

- Create code to find micro satellites

- Investigate creating a web site

3 Project deliverables

Important technologies Working piece of software Documentation

Annotated Bibliography

- [1] W. Press *et al.*, *Numerical recipes in C*. Cambridge University Press Cambridge, 1992, pp. 349–361, 0123456789.

This is my annotation. I can add in comments that are in **bold** and *italics and then other content*.

- [2] M. Neal, J. Feyereisl, R. Rascunà, and X. Wang, “Don’t touch me, I’m fine: Robot autonomy using an artificial innate immune system,” in *Proceedings of the 5th International Conference on Artificial Immune Systems*. Springer, 2006, pp. 349–361.

This paper...

- [3] H. M. Dee and D. C. Hogg, “Navigational strategies in behaviour modelling,” *Artificial Intelligence*, vol. 173(2), pp. 329–342, 2009.

This is my annotation. I should add in a description here.

- [4] Various, “Fail blog,” <http://www.failblog.org/>, Aug. 2011, accessed August 2011.

This is my annotation. I should add in a description here. A longer piece of text.

- [5] S. Duckworth, “A picture of a kitten at Hellifield Peel,” <http://www.geograph.org.uk/photo/640959>, 2007, copyright Sylvia Duckworth and licensed for reuse under a Creative Commons Attribution-Share Alike 2.0 Generic Licence. Accessed August 2011.

This is my annotation. I should add in a description here.