Innovation Report

Your Innovation Report will give judges a better idea of your innovation's background information and your methods. It requires that you select only what is important and state that in a clear and concise way.

Report Requirements:

1. The only identifying information should be the title of your project. Do not include your name or contact information.
2. Make sure the format and your writing is clear and concise, with correct spelling and grammar. Use either Times New Roman or Arial fonts, size 12, and 1.5 line spacing.
3. Do not include raw data, detailed observations or any appendices in your report. The research report is a brief summary and should not include the main materials.
4. Limit your report to one page.
5. Your report must be written in your own words.

What To Include in Your Summary Report:

Introduction & Background:
Describe why you chose to do your project. What made you think of this idea?

Purpose:
Describe the more specific objectives of the project. What are you trying to accomplish?

Design Criteria:
Describe in detail how you know if your innovation is successful.

Procedure or Methodology:
Write a brief outline of the materials and methods used in the development of your innovation, not a detailed description.

Results:
Summarize what you found and show how that relates to the Purpose. A brief discussion of the limitations, or suggestions for further research, may be included.

Conclusions & Recommendations:
Briefly answer the problem posed in the Purpose section.

Acknowledgments:
Recognition should be given to all who provided significant assistance to the researcher in the development of the project.

Sample Summary Report

A Novel Algorithm for Identifying Sequence Motifs

Introduction and Background

DNA (deoxyribonucleic acid) consists of monomers (nucleotides) with three key components: a sugar, a phosphate group, and a nitrogenous base (adenine, cytosine, guanine, or thymine). DNA is a hereditary material in all forms of life. It carries genetic instructions used for growth, development, and reproduction. Motifs are short sequence patterns that represent the fundamental units of biological function, and they can encode protein, or facilitate DNA and RNA interactions, as well as catalytic functions. They are found in both the coding and non-coding parts of the genome. Motifs allow for the identification of genes that contribute to specific functions of an organism. Motif discovery involves the identifying of motifs within sequences through experimental or computational means. Finding these motifs experimentally is very expensive in terms of time and resources, which is why many algorithms have been developed to identify motifs through computational means.

Purpose

The purpose of this project is to develop a novel algorithm that will attempt to solve the problem of finding DNA sequence motifs in an original way to increase the efficiency and accuracy of motif discovery.

Design Criteria

A successful algorithm is one that can accurately identify motifs within known sequences, i.e. the motifs have already been experimentally determined. The algorithm should identify these motifs faster than current algorithms such as the MEME (Multiple Expectation-maximization for Motif Elicitation) or GLAM2 (Gapped local alignment of motifs).

Procedure

The first step in this project was to read the large data files containing sequences into a nice data structure for manipulating in Python. After reading the data, from files in FASTA format, into Python, I explored different algorithms to identify the most prominent motifs. I recorded the number of times a k-mer (sequence of length k) appeared in the set of sequences and used it as the primary value for ranking motifs. However, it became apparent that analyzing only one sequence for motifs was not returning accurate motifs. Next, I randomly shuffled the data set obtained from a cell to create another data set. Then, I was able to compare these two sets of sequences to determine which motifs were unique to the original sequence by using the ratio of the frequencies of the motifs in the two different sequences. After this, I calculated the p-value (using the binomial test) to test the statistical significance of motif occurrences and rank motifs more accurately.

Results

When analyzing the PITX1 (paired-like homeodomain 1) sequences, which were obtained experimentally, the target motif was the “GGATTA” binding motif. I found it to be prominent but not the top ranked motif. However, I found many variants of the “CAGCTG” motif to have very high rankings. When I further researched the cause of this abnormality, I discovered that the “CAGCTG” motif is an E-box (enhancer-box) DNA response element that acts as a protein-binding site. Previous studies have experimentally confirmed that the E-box and PITX1 motifs co-occur due to protein-protein interactions between the proteins encoded by the genes containing these motifs.

Conclusion

In conclusion, I successfully created a newly designed algorithm for DNA motif discovery that was relatively fast when compared to more complex algorithms such as MEME, since analyzing data sets of over 10,000 sequences took around five minutes. Such a method of DNA motif discovery could help find new motifs and connect them to their biological function.

Acknowledgements

I would like to acknowledge my mentor, <name>, for his support and help throughout the making of this project.