Universiti Teknologi MARA

Mining the DNA Sequence for Race Classification

Siti Nurihan Hj.Ariffin

Thesis submitted in fulfillment of the requirements for
Bachelor of Science (Hons) Information System Engineering
Faculty of Information Technology And
Quantitative Science

May 2008

ACKNOWLEDGEMENT

In the name of Allah, The Most Gracious, The Most Merciful and Him alone is worthy of all praise. Thank you the Almighty God for giving me the strength and ability towards completing my final year project.

First and foremost, I would like to thank to my supervisor Pn. Fauziah Redzuan, for her patient, advice and criticism for betterment to this research. This project wasn't made possible without her full support.

A very special thanks to Pn. Jamaliah bt. Taslim and PM Rashidah bt.Rawi, who has never hesitate in helping me upon the completion of my research .With their guidance and ideas, they help me lots in doing my research.

I also would like to thank to all my friends for their dedication to this research and their never ending passion. And also to my entire respondent that give their cooperation towards my research.

A millions thanks to my parents, especially to my mother Jamaliah@Samsiah bt. Mat who are really understand, encouraging and giving full of her support to me to do my research and my study.

Last but not least, thanks to all those who are assisting me directly or indirectly until I complete my research. Thank you very much for all your support.

Thank you.

ABSTRACT

Bioinformatics is a new field that arises in Malaysia. It offers many benefits especially in the genes and genomic research. By using the computers technology, it helps the researchers to understand the large amount of data in the bioinformatics. By implementing data mining techniques, the researchers are able to analyze the data from different perspective and summarizing the data into useful information. This research focusing on identifying the data mining methods/ techniques for the race classification and predict the races based on their DNA sequence using the appropriate technique. The comparison between classification techniques was done in order to get the best technique for classification. By implementing the existing tool of neural network, the race classification was achieved. From this research, it helps in identifying the techniques for classification.

TABLE OF CONTENTS

TITLE **PAGE**

CHAPTER THREE: RESEARCH APPROACH AND METHODOLOGY

CHAPTER FOUR: ANALYSIS AND FINDINGS