UNIVERSITI TEKNOLOGI MARA

ANALYSIS OF SEQUENCED GENOMIC DNA SAMPLES OF *E. COLI* AND *P. MIRABILIS* USING CLC GENOMICS WORKBENCH

HAFIZUDDIN SHAH BIN SALAMAN

Dissertation submitted in partial fulfillment of the requirements for the bachelor in pharmacy

Faculty of Pharmacy

NOVEMBER 2009

ACKNOWLEDGEMENT

First of all thanks to Allah S.W.T and may His peace and blessings be upon all his prophets for granting me the chance and the ability to successfully complete this study. Then, I would like to express my gratitude to my supervisor Miss Fazleen Haslinda binti Mohd Hatta, for her continuous support in completing this research project. Miss Fazleen has always listened and gave advice in preparation in this project. Her understanding, encouraging and personal guidance have provided a good spirit to me to complete this course. Then, my thank you goes to Associate Professor Dr Teh Lay Kek in helping me while doing the project. Lastly, special thanks to all pre-graduate students and all my friends for their great cooperation and willingness.

TABLE OF CONTENTS

ABSTRACT

The main goal of this study is to analyze sequenced samples of *Escherichia coli* and *Proteus mirabilis* using a program called CLC Genomics Workbench. The program can be used to compare between the sample data collected and reference sequence at Genbank. Further interpretation using reference assembly function in the CLC Genomic Workbench is needed to confirm if the sample is a new strain or of a previously identified species.

CHAPTER ONE

INTRODUCTION

1.1 Sequence technology

James Watson and Francis Crick discovered the double helix in 1953, the twisted-ladder structure of deoxyribonucleic acid (DNA), marking a milestone in the history of science and gave rise to modern molecular biology, which is largely concerned with understanding how genes control the chemical processes within cells. Their discovery yielded ground-breaking insights into the genetic code and protein synthesis. During the 1970s and 1980s, it helped to produce new and powerful scientific techniques, specifically recombinant DNA research, genetic engineering, rapid gene sequencing, and monoclonal antibodies. Subsequently, the Sanger method was developed by Frederick Sanger in 1975. Most DNA sequencing that occurs in medical and research laboratories are performed using sequencers employing variations of the Sanger method. Termed the chain-termination method, it involves a reaction where chain-terminator nucleotides are labeled with fluorescent dyes, combined with fragmented DNA, DNA sequencing primers and DNA polymerase. Each nucleotide in the DNA sequence is labeled with a different dye color and a chromatogram is produced, with each color representing a different letter in the DNA code – A,T,C, or G(Nyren et al, 2006). A new generation of non-Sanger-based sequencing technologies has delivered on its promise of sequencing