

Table of Contents

| | |
|--|------------|
| Abstract | I |
| Acknowledgements..... | I |
| Table of Contents | III |
| List of Figures..... | V |
| Abbreviations | VII |
| Chapter 1 Introduction | 1 |
| 1.1 Introduction..... | 1 |
| 1.1.1 Introduction to Bioinformatics..... | 3 |
| 1.1.2 Bioinformatics: From the Pre-genomics to the Post-Genomics Era | 5 |
| 1.1.3 The Human Genome Project (HGP) | 6 |
| 1.2 Statement of Problems..... | 7 |
| 1.3 Project Objectives | 8 |
| 1.4 Motivation and Contribution | 8 |
| 1.5 Project Organization..... | 9 |
| Chapter 2 Literature Review | 11 |
| 2.1 Web-based Bioinformatics Resources..... | 12 |
| 2.2 National Center for Biotechnology Information (NCBI) | 13 |
| 2.2.1 BLAST | 15 |
| 2.2.2 Entrez..... | 15 |
| 2.2.3 OMIM..... | 16 |
| 2.2.4 PubMed..... | 17 |
| 2.3 European Bioinformatics Institute (EBI)..... | 18 |
| 2.3.1 Databases at EBI | 19 |
| 2.3.2 EMBL - The Nucleotide Sequence Database | 20 |
| 2.3.3 The SwissProt Database | 21 |
| 2.4 GenomeNet- Japanese Bioinformatics Servers..... | 21 |
| 2.4.1 DBGET/LinkDB Integrated Database Retrieval System | 23 |
| 2.4.2 KEGG (Kyoto Encyclopedia of Genes and Genomes) | 24 |
| 2.5 Chapter Summary..... | 24 |

| | |
|--|-----------|
| Chapter 3 Data Mining Using BLAST | 24 |
| 3.1 Introduction to BLAST: Basic Local Alignment Search Tool | 26 |
| 3.1.1 Differences of BLAST programs..... | 28 |
| 3.1.2 BLASTn | 30 |
| 3.1.3 Databases | 30 |
| 3.3 Database Sequence Searching..... | 32 |
| 3.4 Query Sequence Entry | 34 |
| 3.5 Chapter Summary..... | 36 |
| Chapter 4 Methodology and Results | 37 |
| 4.1 Introduction to Concept of Molecular Biology..... | 38 |
| 4.2 About DNA Sequencing..... | 41 |
| 4.3 Submitting a BLAST search step by step..... | 42 |
| 4.3.1 Steps in submitting a Query..... | 42 |
| 4.3.2 Getting the Results | 45 |
| 4.4 BLAST Output Results Analysis | 47 |
| 4.4.1 Results | 48 |
| 4.4.2 Graphic Representation | 50 |
| 4.4.3 Analysis Results of Sample 1 | 51 |
| 4.4.4 Analysis Results of CF1 | 54 |
| 4.5 Chapter Summary..... | 60 |
| Chapter 5 Discussion and Conclusion | 61 |
| 5.1 Discussion..... | 61 |
| 5.2 Ethical Issues in Bioinformatics | 63 |
| 5.3 Future Directions..... | 64 |
| 5.4 Conclusion | 64 |
| References..... | 65 |
| Glossary | 69 |
| Appendix A | 72 |
| Appendix B | 78 |

List of Figures

| | |
|---|----|
| Figure 1.1 Overview of Chapter 1..... | 2 |
| Figure 1.2 Interaction of disciplines that have contributed to the formation of Bioinformatics | 3 |
| Figure 2.1 Overview of Chapter 2..... | 11 |
| Figure 2.2 Primary Public Domain Bioinformatics Servers..... | 12 |
| Figure 2.3 National Center for Biotechnology Information (NCBI) | 14 |
| Figure 2.4 Entrez database Browser System | 16 |
| Figure 2.5 European Bioinformatics Institute..... | 19 |
| Figure 2.6 GenomeNet Japan..... | 23 |
| Figure 3.1 Overview of Chapter 3..... | 25 |
| Figure 3.2 A Basic BLAST Interface..... | 27 |
| Figure 3.3 Overview of BLAST programs | 29 |
| Figure 3.4 BLASTn diagram | 30 |
| Figure 3.5 BLAST sequence submitting Window | 35 |
| Figure 4.1 Overview of Chapter 4..... | 37 |
| Figure 4.2 The MLH1 gene, which is associated with colon cancer in humans, is used in this example..... | 40 |
| Figure 4.3 Sample 1 query sequence is pasted into the search box in BLAST interface.. | 43 |
| Figure 4.4 The interface show the output..... | 46 |
| Figure 4.5 Interface show results of BLAST in process | 47 |
| Figure 4.6 Interface show results obtained for the Query Sample 1 | 48 |
| Figure 4.7 Interface show Results obtained for query sample 1..... | 50 |
| Figure 4.8 Distribution of 9 Blast Hits on the Query Sequence Sample 1 | 51 |
| Figure 4.9 Interface show results obtained for the Query CF1 known sequence | 54 |
| Figure 4.10 Interface show results obtained for query CF1 | 55 |
| Figure 4.11 Distribution of 136 Blast Hits for the Query CF1 | 56 |

List of Table

| | |
|---|----|
| Table 4.1 20 Types of Amino acids | 39 |
|---|----|

Abbreviations

| | |
|--------------|---|
| BLAST | Basic Local Alignment Search Tool |
| DDBJ | DNA Data Bank Of Japan |
| DNA | Deoxyribonucleic acid |
| EBI | European Bioinformatics Institute |
| EMBL | European Molecular Biology Laboratory |
| EST | Expressed Sequence Tag |
| FTP | File Transfer Protocol |
| HGP | Human Genome Project |
| HGST | High Throughput Genomics Sequences |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |
| NCBI | National Center for Biotechnology Information |
| NIH | National Institute of Health |
| NLM | National Library of Medicine |
| OMIM | Online Mendelian Inheritance in Man |
| PDB | Protein Data Bank |
| RNA | Ribonucleic acid |
| STS | Sequence Tagged Sites |