## LIST OF FIGURES

<table>
<thead>
<tr>
<th>No.</th>
<th>CAPTION</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Biology Research</td>
<td>3</td>
</tr>
<tr>
<td>1.2</td>
<td>Partial sequence of the Hepatitis B virus genome retrieved from the GenBank database in FASTA format</td>
<td>4</td>
</tr>
<tr>
<td>1.3</td>
<td>Partial sequence H1N1 Strain Protein sample</td>
<td>5</td>
</tr>
<tr>
<td>1.4</td>
<td>Central Dogma of Biology</td>
<td>7</td>
</tr>
<tr>
<td>1.5</td>
<td>Relationship Between Genes and Proteins</td>
<td>7</td>
</tr>
<tr>
<td>1.6</td>
<td>Information flow in Organisms</td>
<td>9</td>
</tr>
<tr>
<td>1.7</td>
<td>Assumption for sequence alignment</td>
<td>12</td>
</tr>
<tr>
<td>1.8</td>
<td>Aligned sequences samples</td>
<td>12</td>
</tr>
<tr>
<td>2.1</td>
<td>Alignment of two sequences</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>Distinction between Global and Local alignment of two sequences</td>
<td>21</td>
</tr>
<tr>
<td>2.3</td>
<td>Evolutionary Relationship between two similar sequences</td>
<td>22</td>
</tr>
<tr>
<td>2.4</td>
<td>Dot Matrix Pairwise Alignment of two DNA Sequences (G—top) and (H—side). Diagonally sequences of dots indicate areas of contiguous sequences of aligned pairs.</td>
<td>25</td>
</tr>
<tr>
<td>2.5</td>
<td>Dot Matrix Pairwise Alignment of Sequences (G—top) and (H—side) using filter, with a window of 2 and stringency of 1, emphasizes contiguous aligned sequence pairs.</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>Dot matrix comparison of two DNA sequences, AGCTAGGA and GACTAGGC. The diagonal of dots reveals a run of characters CTAGG in the two sequences.</td>
<td>26</td>
</tr>
</tbody>
</table>
2. 7. Score matrix between two sequences GAATTAGTTA and GGATCGA using dynamic programming ...................................................... 27
2. 8. FASTA Algorithm Flowchart ............................................................ 33
2. 9. Hash Table for FASTA. The possible words are keyed to index numbers (right), which are used to represent words in the hash table .............. 33
2. 10. Process outline of MAFFT alignment ................................................ 39
3. 1. CPU Time with DATASET-1 ............................................................... 65
3. 2. Memory used with DATASET-1 ........................................................... 65
3. 3. EFPLCS Execution statistics on Dataset 2- a sample ......................... 66
3. 4. FASTLCS Execution Statistics on Dataset2 – a sample ..................... 67
3. 5. CPU Time with Dataset-2 ................................................................. 67
3. 6. Memory Usage with Dataset-2 ........................................................... 68
4. 1. Effect of h-value on execution time requirement .................................. 78
4. 2. Effect of h-value on memory requirement .......................................... 78
4. 3. Comparison of Memory Utilisation of SRLCS Vs EFPLCS ................. 86
4. 4. Comparison of Execution Time of SRLCS Vs EFPLCS ..................... 86
5. 1. CLUSTAL-W Screen shot ................................................................. 91
5. 2. Sample input file for MUSCLE execution ......................................... 93
5. 3. Sample output after MUSCLE execution (.ln format) ......................... 93
5. 4. Sample Command line dialogue – MUSCLE execution ..................... 94
5. 5. Screen shot of SSEARCH execution ................................................. 96
5. 6. Sample output file on execution of SSEARCH35 ............................ 97
5. 7. Pairwise Identity vs. LCS identification of SRLCS, CLUSTAL-W and MUSCLE ................................................................. 100
5. 8. Precision of identified LCS by SRLCS, CLUSTAL-W and MUSCLE .......... 104
6. 1. Identity Vs Length of LCS ......................................................... 109
6. 2. Scatter Chart for Predicted Vs.Actual LCS With reference to |X Sequence|..... 112
6. 3. Scatter Chart for Predicted Vs.Actual LCS With reference to |Y Sequence|..... 112
6. 4. Scatter Chart for Predicted Vs.Actual LCS with reference to Identity Percentage ........................................................................................................... 112
6. 5. Scatter Chart for Predicted Vs.Actual LCS with reference to Similarity Percentage ........................................................................................................... 112
7. 1. Identification of Membership Function of a target sequence......................... 120
7. 2. Identification strain segments PB2, PB1, HA and NA of H1N1 strain A/Pune/NIV8489/2009 in other samples ............................................................ 124
7. 3. Membership of other HA segment samples in gb:JN600356 HA protein segment ............................................................................................................ 125
8. 1. Structural alignment of goose lysozyme (PDB code 153L), chicken egg white lysozyme (3LZT), and lysozymes from E. coli bacteriophages λ (1AM7) and T4 (1L92) .......................................................................................................... 130
8. 2. Identification of protein family of an unknown protein sequence................. 133
8. 3. Protein sequences comparison with a known specimen ................................. 134
8. 4. Membership Calculator .................................................................................. 135
8. 5. Homology extent calculation of specimen sequence with reference to template sequences ............................................................................................... 136
8. 6. Membership information of Myoglobin HS22-37-PE42 Vs. other organisms ... 139
8. 7. Membership information of Lysozyme HS 10-33PE14 Vs. other organisms..... 141
8. 8. Membership information of Lysosyme ECO81_1_PE245 vs. other organism... 142
8. 9. Membership information of QNR HS 6-36 Vs. other organisms ................... 143