

List of Figures

FIGURE 1. SCOPE OF RESEARCH WORKVIII

FIGURE 2. VARYING GENOME SIZES OF DIFFERENT ORGANISMS3

FIGURE 3. CLIENT SERVER MODEL IMPLEMENTED FOR DISTRIBUTED APPLICATION20

FIGURE 4. TECHNOLOGIES USED TO IMPLEMENT THE DISTRIBUTED APPLICATION29

FIGURE 5. USER INTERFACE WITH THE DISTRIBUTED WEB APPLICATION, TO UPLOAD THE FASTA FILE CONTAINING DNA SEQUENCING DATA.....30

FIGURE 6. WEBPAGE DISPLAYING THE RESULT OF THE QUERY FIRED ON THE DATABASE RESIDING ON THE REMOTE MACHINE31

FIGURE 7. DIFFERENCE BETWEEN THE STRUCTURE OF RNA AND DNA47

FIGURE 8. DNA, AS A CHAIN OF SMALLER MOLECULES OR NUCLEOTIDE BASES.....50

FIGURE 9. THREE DIMENSIONAL VIEW OF NUCLEOTIDE SEQUENCE.....52

FIGURE 10. ONE DIMENSIONAL VIEW OF NUCLEOTIDE SEQUENCE52

FIGURE 11. REPEATED DNA SEQUENCES IN EUKARYOTIC GENOMES61

FIGURE 12. SIGNAL REPRESENTATION OF THE FUNCTION STATED IN EXAMPLE 1 - A STATIONARY SIGNAL64

FIGURE 13. SIGNAL REPRESENTATION OF THE FUNCTION STATED IN EXAMPLE 2 - A NON-STATIONARY SIGNAL64

FIGURE 14. DISCRETE-TIME SIGNAL REPRESENTATION OF THE FUNCTION STATED IN EXAMPLE 166

FIGURE 15. DISCRETE-TIME SIGNAL REPRESENTATION OF THE FUNCTION STATED IN EXAMPLE 266

FIGURE 16. VARIOUS DOMAINS IN WHICH SIGNALS ARE REPRESENTED74

FIGURE 17. WAVELET TRANSFORM (ANALYSIS PHASE)86

FIGURE 18. “MALLAT-TREE” FOR THREE LEVEL WAVELET DECOMPOSITION87

FIGURE 19. INVERSE WAVELET TRANSFORM (SYNTHESIS PHASE)98

FIGURE 20. “MALLAT TREE” FOR SYNTHESIS PHASE.....99

FIGURE 21. DNA SEQUENCING PER DOLLAR IS INCREASING FASTER THAN STORAGE CAPACITY PER DOLLAR.....106

FIGURE 22. VISUAL REPRESENTATION OF FOUR LEVEL DECOMPOSITIONS DISPLAYING REDUCING LENGTH TREND, OF A SEQUENCE AS EXPLAINED IN EXAMPLE 1.....118

FIGURE 23. GRAPH REPRESENTING DATA REDUCTION UPTO NEARLY 64 TIMES FOR METAGENOMIC SEQUENCE WITH ACCESSION No. SRR00675.1.2119

FIGURE 24. GRAPH REPRESENTING DATA REDUCTION UPTO NEARLY 64 TIMES FOR CONTIG 46 OF E.COLI K12 STRAIN.....120

FIGURE 25. GRAPH REPRESENTING DATA REDUCTION UPTO NEARLY 64 TIMES FOR CHROMOSOME 26 OF HOMO SAPIENS121

FIGURE 26. GRAPH REPRESENTING DATA REDUCTION UPTO NEARLY 64 TIMES FOR CHROMOSOME 3 OF C. ELEGANS 122

FIGURE 27. VISUAL REPRESENTATION OF COMPRESSED FORM METAGENOMICS SEQUENCE HAVING ACCESSION NUMBER >GNL|SRA|SRR000675.1.2 EXHS9OF01EH7NX.2, UPTO 64 TIMES REDUCTION, USING PROPOSED ALGORITHM APPLYING MULTI-LEVEL HAAR WAVELET TRANSFORM 123

FIGURE 28. STEPS TO IDENTIFY THE IDENTICAL READS..... 129

FIGURE 29. LIST OF READ NUMBERS OF IDENTICAL READS (STARTING READ NUMBER IS 62)..... 135

FIGURE 30. MATLAB OUTPUT OF SEQUENCE NUMBERS WHICH ARE DUPLICATE READS FOR A SEQUENCING DATA WITH ACCESSION NUMBER SRR065619 136

FIGURE 31. STATISTICAL OUTPUT GENERATED BY MATLAB PROGRAM FOR FINDING DUPLICATE READS. 137

FIGURE 32. MATLAB PROGRAM OUTPUT DEPICTING THE TOTAL NUMBER OF UNIQUE DUPLICATE READS 138

FIGURE 33. IMPROVISED ALGORITHM FOR RECOGNIZING IDENTICAL READS 139

FIGURE 34. VISUAL REPRESENTATION OF SHORT TANDEM REPEAT REGION IN SEQUENCES AS IN EXAMPLE 1, USING MULTI-LEVEL HAAR WAVELET TRANSFORM. 155

FIGURE 35. DETAILED GRAPHICAL REPRESENTATION OF SHORT TANDEM REPEAT REGION IN SEQUENCES AS IN EXAMPLE 1(A), USING MULTI-LEVEL HAAR WAVELET TRANSFORM. 156

FIGURE 36. GRAPHICAL REPRESENTATION OF SHORT TANDEM REPEAT REGION. RED COLOURED ELLIPSE DEFINES THE STR REGION CONTAINING MONO-HOMOPOLYMER 156

FIGURE 37. DETAILED GRAPHICAL REPRESENTATION OF SHORT TANDEM REPEAT REGION IN SEQUENCES AS IN EXAMPLE 1(B), USING MULTI-LEVEL HAAR WAVELET TRANSFORM. 157

FIGURE 38. REPRESENTING THE DISCRETE WAVELET TRANSFORM OF THE SEQUENCE IN EXAMPLE-2 WITH FOUR REPEAT REGIONS. 158

FIGURE 39. THE GRAPHICAL REPRESENTATION OF OUTPUT OF WAVELET-BASED TECHNIQUE FOR SEARCHING 159

FIGURE 40. THE OUTPUT FROM MISA SOFTWARE, REPEATS FOUND WITH DEFAULT PARAMETERS. 160

FIGURE 41. GRAPHICAL REPRESENTATION OF SHORT TANDEM REPEAT REGIONS FROM SEQUENCE EXTRACTED FROM THE THE HUMAN GENOME REFSEQ DATABASE (RELEASE 28). 161