List of Tables

TABLE 1. TABLE STRUCTURE FOR STORING THE SPECIES DETAILS, FOR THE SPECIES OR ORGANISMS, WHOSE DNA SEQUENCING IS PERFORMED 23
TABLE 2. PARTIAL SNAPSHOT OF THE TABLE STRUCTURE FOR STORING THE DETAILS OF THE RUNS OR DNA SEQUENCING EXPERIMENTS, PERFORMED FOR THE SPECIES 23
TABLE 3. TABLE STRUCTURE FOR STORING THE DETAILS OF THE READS GENERATED FROM DNA SEQUENCING OF THE GIVEN SPECIES AND DURING A GIVEN RUN 23
TABLE 4. BRIEF DESCRIPTION OF JAVA PROGRAMS 37
TABLE 5. REPRESENTATION OF COMPUTATIONS OF HAAR WAVELET TRANSFORM 94
TABLE 6. REPRESENTATION OF OF HAAR WAVELET TRANSFORM FOR XN 102
TABLE 7. REPRESENTS THE REDUCTION RATIO OF ORIGINAL SEQUENCE AND COMPRESSED SEQUENCE AFTER BINARY REPRESENTATION AND AFTER FOUR LEVELS OF WAVELET TRANSFORMS, AS PROPOSED IN THE ALGORITHM 125
TABLE 8. REPRESENTS COMPUTATIONAL TIME FOR THE PROPOSED ALGORITHM OF DATA REDUCTION USING WAVELET TRANSFORMS 126
TABLE 9. EIIP VALUES FOR SINGLE INDICATOR OF SEQUENCE REPRESENTATION 131
TABLE 10. RESULT SHOWING NUMBER AND PERCENTAGE OF IDENTICAL READS RECOGNIZED USING THE WAVELET TRANSFORMS BASED ALGORITHM FROM VARIOUS STRAINS OF BACILLUS 132
TABLE 11. TIME TAKEN TO FIND THE IDENTICAL READS USING WAVELET TRANSFORMS BASED ALGORITHM 133
TABLE 12. THE SUBSET OF RECORDS OF THE RESULT GENERATED BY MATLAB PROGRAM USING WAVELET TRANSFORMS ALGORITHM, WHICH REPRESENTS THE READ NOS. OF ALL IDENTICAL READS FOUND IN THE BACILLUS WITH SRA REFERENCE ID. SRR149222 133
TABLE 13. COUNT OF REDUNDANT READS AND THE NO. OF REDUNDANT BASE PAIRS, FOR THE SUBSET OF RECORDS OF READS FOUND IN SRR149222 134
TABLE 14. RESULT REPRESENTING NUMBER AND PERCENTAGE OF ARTIFICIAL DUPLICATE READS WITH 100% IDENTITIES, RECOGNIZED USING THE WAVELET TRANSFORMS BASED ALGORITHM 140
TABLE 15. REPRESENTING DETAILS OF REDUNDANT READS (AFTER RETAINING ONE COPY EACH) IN TERMS OF NUMBER AND PERCENTAGE OF BASE PAIRS 141
TABLE 16. REPRESENTING BYTE CONSERVATION, USING PROPOSED ALGORITHM 141
TABLE 17. REPRESENTING THE SRA READ NO., WITH HIGHEST NUMBER OF DUPLICATE READS IN A GIVEN SRA READ, ALONG WITH LENGTH AND MAXIMUM NUMBER OF COPIES (SRA NO. OF READ AS SPECIFIED IN COLUMN (D)) 142

TABLE 18. REPRESENTING COMPUTATIONAL TIME FOR IDENTIFYING ARTIFICIAL DUPLICATE READS WITH 100% IDENTITIES 143
TABLE 19. RESULT DISPLAYING COMPUTATIONAL TIME FOR IDENTIFYING ARTIFICIAL DUPLICATE READS WITH 100% IDENTITIES RECOGNIZED USING THE WAVELET TRANSFORMS BASED ALGORITHM USED FOR DATA REDUCTION, PRIOR TO COMPARING THE SEQUENCES FOR FINDING EXACT OR NEAR EXACT MATCH 143
TABLE 20. TERMS AND NOTATIONS USED TO DESCRIBE STRS / MICROSATELLITES 146
TABLE 21. REPRESENTATION OF DIPOLE MOMENTS VALUE FOR NUCLEOTIDE BASES . 150
TABLE 22. LIST OF SAMPLE DATA TAKEN FROM HUMAN GENOME REFSEQ (RELEASE 8) FOR REPRESENTING SHORT TANDEM REPEAT REGIONS 162
TABLE 23. BRIEF DESCRIPTION OF MATLAB PROGRAMS 164