

CONTENTS

S. No.	DESCRIPTION	PAGE No.
1	Introduction	1 – 4
2	Review of literature	5 – 29
2.1	Infectious diseases	5
2.2	<i>Streptococcus pneumoniae</i>	6
2.2.1	General characteristics	6
2.2.2	Genetics	6
2.2.3	Cultural characteristics	7
2.2.4	Virulence factors	8
2.2.5	Pathogenesis	9
2.2.6	Laboratory diagnosis	10
2.3	<i>Haemophilus influenzae</i>	11
2.3.1	History	11
2.3.2	Genetics	11
2.3.3	Cultural characteristics	12
2.3.4	Virulence factors	13
2.3.5	Pathogenesis	14
2.3.6	Laboratory diagnosis	14
2.4	Melioidosis	16
2.4.1	History	16
2.4.2	Genome plasticity	17
2.4.3	Virulence factors	17
2.4.4	Pathogenesis	18
2.4.5	Clinical description	19
2.4.6	Laboratory diagnosis	20

S. No.	DESCRIPTION	PAGE No.
2.4.7	Epidemiology	21
2.4.8	Treatment	22
2.5	Ebola virus disease	22
2.5.1	Background	22
2.5.2	Virulent proteins & reservoirs	24
2.5.3	Epidemiology	24
2.5.4	Pathogenesis	25
2.5.5	Laboratory diagnosis	27
2.6	Infectious diseases databases	28
2.6.1	European hepatitis C Virus Database	28
2.6.2	EchoBASE	28
2.6.3	TubercuList	28
2.6.4	<i>Pseudomonas</i> genome database	29
3	Materials and Methods	30 – 42
3.1	Databases design and implementation	30
3.2	Retrieval of data	30
3.2.1	National Center for Biotechnology Information genome database	30
3.2.2	Universal Protein resource	36
3.2.3	PUBMED database	37
3.2.4	Protein Data Bank	37
3.2.5	Genome atlas database	38
3.3	Implementation of tools	39
3.3.1	BLAST	39
3.3.1.1	BLAST algorithm	40
3.3.1.2	Steps involved in the implementation of BLAST	40

S. No.	DESCRIPTION	PAGE No.
3.3.2	GBrowse	41
3.3.2.1	Steps involved in the implementation of GBrowse in databases	41
3.3.3	JmolApplet	42
3.3.3.1	Steps involved in the implementation of JmolApplet in databases	42
4	Results and discussion	43 – 114
4.1	General features of SPGDB, HIGDB, MDB & EDB	46
4.2	Data included in SPGDB	51
4.3	Data included in HIGDB	52
4.4	Data included in MDB	53
4.5	Data included in EDB	55
4.6	SPGDB case studies	55
4.6.1	Case study 1	55
4.6.2	Case study 2	69
4.7	HIGDB case studies	71
4.7.1	Case study 1	71
4.7.2	Case study 2	77
4.8	MDB case studies	79
4.8.1	Case study 1	79
4.8.2	Case study 2	108
4.9	EDB case studies	109
4.9.1	Case study 1	109
4.9.2	Case study 2	113
5	Conclusions	115
6	References	

LIST OF FIGURES

S. No.	DESCRIPTION	PAGE No.
2.1	Pathogenic route for <i>S. pneumoniae</i> infection	9
2.2	Work flow for characterization and determination of a <i>S. pneumoniae</i> isolate	10
2.3	Work flow for identification and characterization of a <i>H. influenzae</i> isolate	15
2.4	Laboratory test used in diagnosis of EVD	27
3.1	Schematic representation of retrieval of various information and tools deployed in developed databases	31
3.2	Screenshot of part of a gbk file of <i>S. pneumoniae</i> R6 genome	35
4.1	Screenshot of SPGDB homepage	44
4.2	Screenshot of HIGDB homepage	44
4.3	Screenshot of MDB homepage	45
4.4	Screenshot of EDB homepage	45
4.5	Main menus and submenus available in SPGDB, HIGDB, MDB and EDB	46
4.6	Jmol view of 3-D structure of phosphomevalonate kinase of <i>S. pneumoniae</i> (PDB ID: 1K47) in SPGDB	49
4.7	Jmol view of 3-D structure of aspartate-semialdehyde dehydrogenase of <i>H. influenzae</i> (PDB ID: 1Q2X) in HIGDB	49
4.8	Jmol view of 3-D structure of Catalase-peroxidase of <i>B. pseudomallei</i> (PDB ID: 2DV1) in MDB	50
4.9	Jmol view of 3-D structure of RNA binding protein of <i>Zaire ebolavirus</i> (PDB ID: 3L29) in EDB	50
4.10	Number of Com-box in <i>S. pneumoniae</i> 670-6B genome	57
4.11	Number of Com-box in <i>S. pneumoniae</i> 70585 genome	57
4.12	Number of Com-box in <i>S. pneumoniae</i> A026 genome	58
4.13	Number of Com-box in <i>S. pneumoniae</i> AP200 genome	58

S. No.	DESCRIPTION	PAGE No.
4.14	Number of Com-box in <i>S. pneumoniae</i> ATCC 700669 genome	59
4.15	Number of Com-box in <i>S. pneumoniae</i> CGSP14 genome	59
4.16	Number of Com-box in <i>S. pneumoniae</i> D39 genome	60
4.17	Number of Com-box in <i>S. pneumoniae</i> G54 genome	60
4.18	Number of Com-box in <i>S. pneumoniae</i> gamPNI0373 genome	61
4.19	Number of Com-box in <i>S. pneumoniae</i> Hungary 19A-6 genome	61
4.20	Number of Com-box in <i>S. pneumoniae</i> INV104 genome	62
4.21	Number of Com-box in <i>S. pneumoniae</i> INV200 genome	62
4.22	Number of Com-box in <i>S. pneumoniae</i> JJA genome	63
4.23	Number of Com-box in <i>S. pneumoniae</i> OXC141 genome	63
4.24	Number of Com-box in <i>S. pneumoniae</i> P1031 genome	64
4.25	Number of Com-box in <i>S. pneumoniae</i> R6 genome	64
4.26	Number of Com-box in <i>S. pneumoniae</i> SPN034156 genome	65
4.27	Number of Com-box in <i>S. pneumoniae</i> SPN034183 genome	65
4.28	Number of Com-box in <i>S. pneumoniae</i> SPN994038 genome	66
4.29	Number of Com-box in <i>S. pneumoniae</i> SPN994039 genome	66
4.30	Number of Com-box in <i>S. pneumoniae</i> SPNA45 genome	67
4.31	Number of Com-box in <i>S. pneumoniae</i> ST556 genome	67
4.32	Number of Com-box in <i>S. pneumoniae</i> Taiwan 19F-14 genome	68
4.33	Number of Com-box in <i>S. pneumoniae</i> TCH8431/19A genome	68
4.34	Number of Com-box in <i>S. pneumoniae</i> TIGR4 genome	69
4.35	GBrowse visualization of <i>S. pneumoniae</i> TIGR genome from 1,002,689 to 1,072,420	70

S. No.	DESCRIPTION	PAGE No.
4.36	GBrowse visualization of <i>PsaA</i> gene in <i>S. pneumoniae</i> G54	71
4.37	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> 86-028NP genome	72
4.38	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> 10810 genome	73
4.39	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> F3031 genome	73
4.40	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> F3047 genome	74
4.41	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> KR494 genome	74
4.42	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> Rd KW20 genome	75
4.43	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> PittEE genome	75
4.44	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> PittGG genome	76
4.45	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> R2846 genome	76
4.46	Number of motif “A ₂ GTGCGGT” in <i>H. influenzae</i> R2866 genome	77
4.47	GBrowse visualization of <i>H. influenzae</i> Rd KW20 genome from 937,849 to 1,036,694	78
4.48	GBrowse visualization of IgA1 gene of <i>H. influenzae</i> Rd KW20	79
4.49	Number of CpG motif in <i>B. pseudomallei</i> 576, Chromosome 1 genome	82
4.50	Number of CpG motif in <i>B. pseudomallei</i> 576, Chromosome 2 genome	83
4.51	Number of CpG motif in <i>B. pseudomallei</i> 668, Chromosome 1 genome	83

S. No.	DESCRIPTION	PAGE No.
4.52	Number of CpG motif in <i>B. pseudomallei</i> 668, Chromosome 2 genome	84
4.53	Number of CpG motif in <i>B. pseudomallei</i> 1026b, Chromosome 1 genome	84
4.54	Number of CpG motif in <i>B. pseudomallei</i> 1026b, Chromosome 2 genome	85
4.55	Number of CpG motif in <i>B. pseudomallei</i> 1106a, Chromosome 1 genome	85
4.56	Number of CpG motif in <i>B. pseudomallei</i> 1106a, Chromosome 2 genome	86
4.57	Number of CpG motif in <i>B. pseudomallei</i> 1710b, Chromosome 1 genome	86
4.58	Number of CpG motif in <i>B. pseudomallei</i> 1710b, Chromosome 2 genome	87
4.59	Number of CpG motif in <i>B. pseudomallei</i> BDP, Chromosome 1 genome	87
4.60	Number of CpG motif in <i>B. pseudomallei</i> BDP, Chromosome 2 genome	88
4.61	Number of CpG motif in <i>B. pseudomallei</i> BGR, Chromosome 1 genome	88
4.62	Number of CpG motif in <i>B. pseudomallei</i> BGR, Chromosome 2 genome	89
4.63	Number of CpG motif in <i>B. pseudomallei</i> BPC006, Chromosome 1 genome	89
4.64	Number of CpG motif in <i>B. pseudomallei</i> BPC006, Chromosome 2 genome	90
4.65	Number of CpG motif in <i>B. pseudomallei</i> BSR, Chromosome 1 genome	90
4.66	Number of CpG motif in <i>B. pseudomallei</i> BSR, Chromosome 2 genome	91
4.67	Number of CpG motif in <i>B. pseudomallei</i> HB PUB10134a, Chromosome 1 genome	91

S. No.	DESCRIPTION	PAGE No.
4.68	Number of CpG motif in <i>B. pseudomallei</i> HB PUB10134a, Chromosome 2 genome	92
4.69	Number of CpG motif in <i>B. pseudomallei</i> HB PUB10303a, Chromosome 1 genome	92
4.70	Number of CpG motif in <i>B. pseudomallei</i> HB PUB10303a, Chromosome 2 genome	93
4.71	Number of CpG motif in <i>B. pseudomallei</i> K96243, Chromosome 1 genome	93
4.72	Number of CpG motif in <i>B. pseudomallei</i> K96243, Chromosome 2 genome	94
4.73	Number of CpG motif in <i>B. pseudomallei</i> Mahidol 1106a, Chromosome 1 genome	94
4.74	Number of CpG motif in <i>B. pseudomallei</i> Mahidol 1106a, Chromosome 2 genome	95
4.75	Number of CpG motif in <i>B. pseudomallei</i> MSHR146, Chromosome 1 genome	95
4.76	Number of CpG motif in <i>B. pseudomallei</i> MSHR146, Chromosome 2 genome	96
4.77	Number of CpG motif in <i>B. pseudomallei</i> MSHR305, Chromosome 1 genome	96
4.78	Number of CpG motif in <i>B. pseudomallei</i> MSHR305, Chromosome 2 genome	97
4.79	Number of CpG motif in <i>B. pseudomallei</i> MSHR346, Chromosome 1 genome	97
4.80	Number of CpG motif in <i>B. pseudomallei</i> MSHR511, Chromosome 1 genome	98
4.81	Number of CpG motif in <i>B. pseudomallei</i> MSHR511, Chromosome 2 genome	98
4.82	Number of CpG motif in <i>B. pseudomallei</i> MSHR520, Chromosome 1 genome	99
4.83	Number of CpG motif in <i>B. pseudomallei</i> MSHR520, Chromosome 2 genome	99

S. No.	DESCRIPTION	PAGE No.
4.84	Number of CpG motif in <i>B. pseudomallei</i> MSHR1655, Chromosome 1 genome	100
4.85	Number of CpG motif in <i>B. pseudomallei</i> MSHR1655, Chromosome 2 genome	100
4.86	Number of CpG motif in <i>B. pseudomallei</i> MSHR5848, Chromosome 1 genome	101
4.87	Number of CpG motif in <i>B. pseudomallei</i> MSHR5848, Chromosome 2 genome	101
4.88	Number of CpG motif in <i>B. pseudomallei</i> MSHR5855, Chromosome 1 genome	102
4.89	Number of CpG motif in <i>B. pseudomallei</i> MSHR5855, Chromosome 2 genome	102
4.90	Number of CpG motif in <i>B. pseudomallei</i> MSHR5858, Chromosome 1 genome	103
4.91	Number of CpG motif in <i>B. pseudomallei</i> MSHR5858, Chromosome 2 genome	103
4.92	Number of CpG motif in <i>B. pseudomallei</i> MAU20B-16, Chromosome 1 genome	104
4.93	Number of CpG motif in <i>B. pseudomallei</i> NAU20B-16, Chromosome 2 genome	104
4.94	Number of CpG motif in <i>B. pseudomallei</i> NAU35A-3, Chromosome 1 genome	105
4.95	Number of CpG motif in <i>B. pseudomallei</i> NAU35A-3, Chromosome 1 genome	105
4.96	Number of CpG motif in <i>B. pseudomallei</i> NCTC 13179, Chromosome 1 genome	106
4.97	Number of CpG motif in <i>B. pseudomallei</i> NCTC 13179, Chromosome 2 genome	106
4.98	Number of CpG motif in <i>B. pseudomallei</i> NCTC 13178, Chromosome 1 genome	107
4.99	Number of CpG motif in <i>B. pseudomallei</i> NCTC 13178, Chromosome 2 genome	107

S. No.	DESCRIPTION	PAGE No.
4.100	GBrowse visualization of <i>B. pseudomallei</i> K96243, Chromosome 1 from 2,792,230 to 2,799,229	108
4.101	GBrowse visualization of type IV fimbrial pilin protein in <i>B. pseudomallei</i> K96243, Chromosome 1	109
4.102	Number of CpG motif in <i>Bundibugyo ebolavirus</i> genome	110
4.103	Number of CpG motif in <i>Reston ebolavirus</i> genome	111
4.104	Number of CpG motif in <i>Sudan ebolavirus</i> genome	111
4.105	Number of CpG motif in <i>Taiforest ebolavirus</i> genome	112
4.106	Number of CpG motif in <i>Zaire ebolavirus</i> genome	112
4.107	GBrowse visualization of <i>Zaire ebolavirus</i> from 2,000 to 2,399	113
4.108	GBrowse visualization of gp protein of <i>Zaire ebolavirus</i>	114