WHOLE GENOME SEQUENCE CHARACTERIZATION OF PASTEURELLA MULTOCIDA ISOLATED FROM DIFFERENT ANIMAL SPECIES

A THESIS SUBMITTED TO THE ANAND AGRICULTURAL UNIVERSITY IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF

Doctor of Philosophy IN

ANIMAL BIOTECHNOLOGY

BY

VIRAL B. AHIR
M. V. Sc.
(Reg. No. 04-1070-2009/14)

DEPARTMENT OF ANIMAL BIOTECHNOLOGY COLLEGE OF VETERINARY SCIENCE AND A.H. ANAND AGRICULTURAL UNIVERSITY ANAND - 388 001 [GUJARAT]

2015
Dedicated to
My Beloved Family
And GOD
ABSTRACT
ABSTRACT

"WHOLE GENOME SEQUENCE CHARACTERIZATION OF PASTEURELLA MULTOCIDA ISOLATED FROM DIFFERENT ANIMAL SPECIES"

Name of Student Name of Major Advisor
Viral B. Ahir Dr. M. K. Jhala

Department of Animal Biotechnology
College of Veterinary Science and Animal Husbandry
Anand Agricultural University, Anand-388 001
Gujarat, India

Pasteurella multocida is a commensal microorganism of the upper respiratory track of many animal and avian species and is responsible for wide range of diseases in domestic animals and poultry. Despite vaccination of the dairy animals particularly against Haemorrhagic Septicaemia (HS), several outbreaks occur regularly in Gujarat as well as in other parts of India. Whole genome sequencing is a recent advanced approach for understanding of genetic makeup of an organism as well for identification of virulence genes/factors responsible for the disease process in host.

In order to sequence whole genome of P. multocida and to elucidate virulence associated genes, five isolates of P. multocida were sequenced using pyrosequencing based approach of 454 GS FLX Titanium. All the five isolates viz. P52 vaccine strain (P52VAC), poultry (Anand1_poultry), goat (Anand1_goat), buffalo (Anand1_buffalo) and cattle (Anand1_cattle) were identified and characterized based on biochemical and cultural characters and subsequently confirmed by PM-PCR. For sequencing of the whole genome of organisms, dsDNA libraries were prepared for all the five isolates and quantity as well as quality checks were done using Agilent Bioanalyzer as well as TBS fluorometer. dsDNA library of each of the five isolates was amplified using emPCR and positive clonal amplified DNA beads were used for sequencing
after annealing of sequencing primers. After completion of the sequencing run, data generated in form of images were converted into reads using GS Run Browser.

After signal processing, total 118843, 113997, 105729, 134886 and 31346 reads were generated which yielded 42,598,100 (42.59Mb), 29,000,497 (29.00Mb), 21,890,353 (21.89Mb), 39,756,349 (39.75Mb) and 7,429,658 (7.42Mb) of sequence bases for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. Coverage obtained for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle was 18.87, 12.85, 9.70, 17.61 and 3.29 respectively.

All the reads after signal processing were mapped with the reference genome available for a poultry isolate Pm70 at NCBI using GS Reference mapper. Mapping of the isolates P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle resulted in 38,079,806 (89.52%), 20,085,356 (87.38%), 19,867,143 (90.81%), 25,095,466 (63.22%) and 6,145,156 (82.87%) mapped bases with 105327, 97674, 95092, 86765 and 24967 mapped reads. Remaining reads which were not mapped by GS Reference Mapper, were used for de novo assembly using GS De Novo Assembler for finding sequences which code for plasmid of P. multocida. None of the de novo assembled sequences matched to plasmid.

For sequence analysis and finding of virulence associated genes in P. multocida, two different annotation pipelines were used viz. Rapid Annotation using Subsystem Technology (RAST) and Prokaryotic Genome Automatic Annotation Pipeline (PGAAP). For RAST analysis, all the contigs generated after reference mapping with Pm70 uploaded at RAST server. RAST is a subsystem based annotation pipeline which generated 2,273,366bp (2.27Mb), 2,227,943bp (2.22Mb), 2,285,382bp
(2.28Mb), 2,045,610bp (2.04Mb) and 1,438,517bp (1.43Mb) of genome with 209, 489, 349, 2188, and 3152 contigs for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively and 68, 54, 54, 40 and 0 RNA. Based on RAST analysis, highest abundance of subsystem were assigned to ‘amino acids and derivatives’, ‘carbohydrates’, ‘protein metabolism’ and ‘cofactor and vitamins, prosthetic groups and pigments’. As expected, no subsystem was assigned to ‘photosynthesis’ and ‘motility and chemotaxis’ group as Pasteurella is a non-motile organism and is not photosynthetic.

Due to less coverage (3.29X) obtained for the Anand1_cattle isolate, it was omitted from the RAST based comparative analysis. Subsystem based genes/proteins assigned to the other four isolates under ‘virulence, disease and defence’ category ranged from 47 to 54 in number. There were presence of DedA, DedD and toxin under ‘colicin and bacteriocin production’ in P52 vaccine strain, poultry and goat isolates. Genes gyrA, gyrB, Parc and ParD under ‘resistance to fluoroquinolones’ were present in all the four isolates. There was also presence of negative regulator of betalactamase expression, BLR gene leading to resistance expressed by this organism as well as multidrug resistance efflux pump cluster genes, MATE (Multidrug and toxin extrusion), MacA and MacB (Macrolide specific efflux protein) in P52 vaccine strain, poultry and goat isolates.

For PGAAP analysis, all the reads generated after sequencing run were submitted to the PGAAP pipeline of NCBI after removing sequences less than 200bp. PGAAP analysis revealed genome size of 2,273,366bp (2.27Mb), 2,227,943bp (2.22Mb), 2,285,382bp (2.28Mb), 2,045,610bp (2.04Mb) and 1,438,517bp (1.43Mb) with 40.40%, 40.20%, 40.50%, 40.90% and 41.00% of G+C contents for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively.
Total number of coding sequences (CDS) were 2066, 2337, 2319, 3258 and 3623; total number of protein encoding genes (PEG) were 2194, 2284, 2266, 3218 and 3590, and total number of RNA assigned were 64, 53, 53, 41 and 33 for P52VAC, Anandl_poultry, Anandl_goat, Anandl_buffalo and Anandl_cattle, respectively.

Deciphering virulence mechanism is one of the most useful application of bacterial genomics to understand the molecular intricacies involved in disease mechanism as well as for understanding host-pathogen interactions. For this purpose, genes associated with virulence were downloaded from annotation files available at (http://www.ncbi.nlm.nih.gov/genome/genomes/912/) in ‘Protien’ column/section to find out gene locus/id. After manually searching for the virulence associated genes, 55 important genes were selected based on the available literature. These 55 genes grouped under seven broad categories viz. capsule, fimbriae and adhesion, iron metabolism, outer membrane protein, superoxide dismutase, sialic acid metabolism and transcription regulation. Out of these seven categories, all the five genes falling under three categories i.e. SodA and SodC under superoxide dismutase, NanH and NanB under sialic acid metabolism and Fis under transcription regulation category were present in all the five isolates.

Nine genes involved in capsule production were found, out of which, PglA and KmtI were present in all the five isolates, while HyaE was present only in the goat isolate. HexA and HexC genes were absent in buffalo and cattle isolates, while HexB and HexD were absent in goat and cattle isolates. KpsF gene was absent in poultry and cattle isolates. Gene LctP was present only in goat and cattle isolates.

Sixteen genes were found under the category of fimbriae and adhesion, of which, Hsf, PfblB2, PfblR, PfblB, PlpB, and Plp4 genes were found in all the five isolates.
isolates of *P. multocida* studied. *HofC* gene was absent in P52 vaccine strain, whereas *PlpE* gene was absent in cattle as well as buffalo isolates. *ComE* gene was absent in P52 vaccine strain, while *TadE* was absent in buffalo isolate. *PfhB1* and *PlpP* genes were absent in cattle isolate, while *RcpA* and *RcpB* were absent in buffalo isolate. *ClpB* gene was absent in P52 vaccine strain and cattle isolates, whereas *TadF* was absent in buffalo and cattle isolates.

For iron metabolism, 16 genes were found, of which, *ExbB, FbpB, HbpA, HgbA, HemU, OmpW, Rfb* and *RffG* genes were present in all the five isolates studied. Genes *FbpA* and *TonB* were absent in cattle isolate, *TbpA* and *TonB* dependent lactoferrin and transferrin receptor were absent in goat isolate. Gene *FbpC* was absent in buffalo isolate. *TonB*-dependent receptor was present in poultry and goat isolates only. P52 vaccine strain was having a presence of translocation protein *TolB*, whereas *HemR* gene was absent.

For outer membrane proteins, out of nine genes found, *HasR, LppB, LspB, OmpH* and *PtfA* genes were present in all the five isolates of *P. multocida*. Outer membrane protein *LolB* was found in P52 vaccine strain as well as in cattle isolate. *VacJ* gene was absent in goat and cattle isolates. *Oma87* gene was absent only in poultry isolate. Gene *PfhA* was found present only in P52 vaccine strain.

This study is apparently the first attempt in India involving local *P. multocida* isolates from four different species and a vaccine strain for the purpose of identifying virulence genes/virulence associated genes using modern biotechnological tools like pyrosequencing based whole genome sequencing. The study aids in data of whole genome sequencing of bacterial pathogens particularly for *P. multocida* and also provides new insight into their genomic characters and possible molecular
mechanisms involved in disease process. The present findings would provide a much needed base for further screening of virulence associated genes and identification of certain markers for early diagnosis as well as characterization of *P. multocida*, which continues to pose challenges as a menace against the health management of animals. Genes which have been found in all the *P. multocida* isolates under the study can be explored as specific probes for the early diagnosis of the disease. Further, future scientific endeavors targeting the vaccine design for *P. multocida* may get a scientific support from this data, so as to formulate modern and more effective vaccines, for better animal health.
CERTIFICATE

This is to certify that the thesis entitled "WHOLE GENOME SEQUENCE CHARACTERIZATION OF PASTEURELLA MULTOCIDA ISOLATED FROM DIFFERENT ANIMAL SPECIES" submitted by VIRAL B. AHIR (Reg. No. 04-1070-2009/14) in partial fulfillment of the requirements for the award of the degree of DOCTOR OF PHILOSOPHY in the subject of ANIMAL BIOTECHNOLOGY of the Anand Agricultural University is a record of bona fide research work carried out by him under my guidance and supervision and the thesis has not previously formed the basis for the award of any degree, diploma or other similar title.

Place: Anand
Date: 31. 03. 2015

(M.K. Jhala)
MAJOR ADVISOR
CERTIFICATE

This is to certify that, I have no objection for supplying to any scientist only one copy or any part of this thesis at a time through reprographic process, if necessary, for rendering reference services in a library or documentation centre.

(Viral B. Ahir)

Place: Anand
Date: 31. 03. 2015

(M. K. Jhala)
MAJOR ADVISOR
ACKNOWLEDGEMENT
ACKNOWLEDGEMENT

Towards the end of my endeavor, it is the right moment to express my profound regard to all those who have directly or indirectly involved, helped me to accomplish my study and enlightened me with touch of their knowledge and constant encouragement. It is difficult to mention all who have helped me and therefore, I start with expressing my indebtedness to everyone who generously extended their help without any fettering.

First and foremost, I would like to express my deepest gratitude and most sincere thanks to my Major Advisor Dr. M. K. Jhala, Associate Director of Research and former Professor and Head, Department of Veterinary Microbiology, Veterinary College, Anand, for whom, words can never express the high regards and love I have towards him. He has supported me throughout my thesis work with his patience and knowledge, whilst allowing me the room to work in my own way. He provided me invaluable and critical suggestions, scientific acumen, perspicacious remarks, scholarly guidance, active persuasion and supervision throughout the course of my study and research work. Without his support, I could not have completed my work in time.

Words are inadequate for the guidance given by Dr. C. G. Joshi, Professor and Head, Department of Animal Biotechnology, Veterinary College, Anand for his supervision, advice, crucial contribution and guidance from very early stage of this study as well as giving me extraordinary experience throughout the work. He provided me unflinching encouragement and support in various ways. I consider myself fortunate to work under staid and providing me a Senior Research Fellowship during the study under his project. I am indebted to him more than he knows.

I would like to record my sincere gratitude and thanks to my Minor Advisor Dr. Ashish Roy, professor, Veterinary Microbiology, Veterinary College, Anand for introspective guidance with constructive suggestions, deliberative discussions and active persuasion during the period of my study.

I am equally grateful to the members of my Advisory Committee, Dr. D. N. Rank, Professor and Head, Department of Animal Genetics and Breeding, and Dr. A. M. Pandé, Professor and Head, Department of Biochemistry and Biotechnology for their constant support and keen interest in my work during the course of study.

No words of appreciation will ever be sufficient to express my feelings of gratitude to my seniors, colleagues and Assistant Professors Dr. Subhsh Jakhesara, Dr. Prakash Koringa, Dr. Rajyuddin Mathakaya, Dr. Bhard Bhandari and Dr. Umed Ramani for their constant help, motivation and invaluable suggestions during my research work. I am also thankful to my friend and colleague Dr. Ajai Tripathi for his whole hearted co-operation and suggestion during the work.
I am proudly thankful to Dr. J. V. Solanki, Dean and Principal (Retd.) and Dr. A. M. Thaker, Dean and Principal College of Veterinary Science and Animal Husbandry, AUA, Anand, for their generous attitude in providing necessary facilities to carry out the research work.

I am thankful to Director of Animal Husbandry, Krushibhavan, Gandhinagar, Gujarat State to grant me NOC in time to complete my Ph. D. thesis.

I will always be thankful to my junior colleagues Drs. Jitu, Divyesh, Dhaval, Ritesh, Alpesh, Prasan, Tejas, Rajput, Sana, Buhecha, Vala, Bhumi, Aishwarya and Rashmi for their unreserved help and well wishes as sources of constant inspiration for me.

I am also thankful to my colleagues at Ome Research Facility (ORF), Vaibhav, Urvik, Ankit, Tejas, Ramesh, Namrata, Jaina, Manisha, Bhaskar, Vimal, Nidhi, Amit, Ishan, Neelam and Reena for their unreserved help and well wishes to complete my study. I am highly thankful to departmental staff, VaghjiBhai, BijalBhai, Jaysheeben, MacwanBhai, MehtabBhai, Ravjikaka and HarishBhai for keeping friendly environment and their cooperation during the course of study.

Words are just not sufficient to express my feeling for love, affections and encouragement extended by my beloved Father and Mother. Truly they deserve more than a word of thanks, because without their psychological support, higher study for me would have remained a mere dream. Their sacrifice surpasses all materialistic achievements. Besides them, I acknowledge my sisters and aunts for their infinite love, blessings and constant encouragement. I convey my heartfelt gratitude to all my family members especially my uncle late shri D.C. Bharwad for his blessings and bring me up to this stage.

Finally, and most importantly, I would like to thank my wife Rita for her support, encouragement to complete Ph.D. and unwavering love were undeniably the bedrock from the start of this work. She was always there cheering me up and stood by me through the good times and bad. Her tolerance of my occasional worst moods is a testament in itself of her unyielding devotion and love. My special and sincere thanks are also due to my affectionate son Shrey and daughter Mishwafon for their love and whose sacrifices, which were realized by our loss of precious time together, were for me the most painful and humbling of all.

At last but not the least, I thank Almighty GOD for giving me patience and strength to overcome the difficulties, which crossed my way in accomplishment of this endeavor. I know it is always there for me and always helps me to get out of all intriguing situations I get myself into.

Place: Anand
Date: 31. 03. 2015

(Viral B. Ahir)
# CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER NO.</th>
<th>TITLE</th>
<th>PAGE NO.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1-7</td>
</tr>
<tr>
<td>II</td>
<td>REVIEW OF LITERATURE</td>
<td>8-66</td>
</tr>
<tr>
<td>III</td>
<td>MATERIALS AND METHODS</td>
<td>67-100</td>
</tr>
<tr>
<td>IV</td>
<td>RESULTS AND DISCUSSION</td>
<td>101-174</td>
</tr>
<tr>
<td>V</td>
<td>SUMMARY AND CONCLUSIONS</td>
<td>175-182</td>
</tr>
<tr>
<td></td>
<td>REFERENCES</td>
<td>I-LX</td>
</tr>
</tbody>
</table>
**LIST OF PLATES**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Plate No.</th>
<th>Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.1</td>
<td>Agarose gel electrophoresis of extracted genomic DNA of <em>P. multocida</em> isolates</td>
<td>104</td>
</tr>
<tr>
<td>2</td>
<td>4.2</td>
<td>PM PCR based identification of <em>Pasteurella multocida</em> isolates showing 460bp product size for all the isolates.</td>
<td>104</td>
</tr>
<tr>
<td>3</td>
<td>4.3</td>
<td>GS FLX library calculation for all the five isolates of <em>P. multocida</em></td>
<td>106</td>
</tr>
<tr>
<td>4</td>
<td>4.4</td>
<td>Fluogram of nebulized DNA sample after dsDNA library preparation</td>
<td>106</td>
</tr>
<tr>
<td>5</td>
<td>4.5</td>
<td>Clustal multiple alignment of all <em>Pasteurella multocida</em> isolates with Pm70 reference strain for <em>HemU</em> DNA sequence using MUSCLE</td>
<td>107</td>
</tr>
<tr>
<td>6</td>
<td>4.6</td>
<td>Read length distribution of representative sample P52VAC of <em>P. multocida</em></td>
<td>107</td>
</tr>
<tr>
<td>7</td>
<td>4.7</td>
<td>Subsystem distribution of P52VAC isolate of <em>Pasteurella multocida</em> (based on RAST annotation server)</td>
<td>124</td>
</tr>
<tr>
<td>8</td>
<td>4.8</td>
<td>Subsystem distribution of Anand1_poultry isolate of <em>Pasteurella multocida</em> (based on RAST annotation server)</td>
<td>127</td>
</tr>
<tr>
<td>9</td>
<td>4.9</td>
<td>Subsystem distribution of Anand1_goat isolate of <em>Pasteurella multocida</em> (based on RAST annotation server)</td>
<td>127</td>
</tr>
<tr>
<td>10</td>
<td>4.10</td>
<td>Subsystem distribution of Anand1_buffalo isolate of <em>Pasteurella multocida</em> (based on RAST annotation server)</td>
<td>127</td>
</tr>
<tr>
<td>11</td>
<td>4.11</td>
<td>Subsystem distribution of Anand1_cattle isolate of <em>Pasteurella multocida</em> (based on RAST annotation server)</td>
<td>127</td>
</tr>
<tr>
<td>12</td>
<td>4.12</td>
<td>G+C content of the genome for all the five isolate of <em>P. multocida</em></td>
<td>127</td>
</tr>
<tr>
<td>Page</td>
<td>Section</td>
<td>Description</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>4.13</td>
<td>Graph showing number and length of contigs for representative sample (Anand1_goat) isolate of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>4.14</td>
<td>No. of CDS and PEG for all the isolates of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>4.15</td>
<td>Comparison of the five isolates of <em>P. multocida</em> and Pm70 genome sequences using progressive Mauve alignment</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>4.16</td>
<td>Phylogenetic comparison among <em>Pasteurella multocida</em> species whole genome (complete and incomplete) sequences using MEGA6</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>4.17</td>
<td>Circular genome map of P52VAC isolate of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4.18</td>
<td>Circular genome map of Anand1_poultry isolate of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>4.19</td>
<td>Circular genome map of Anand1_goat isolate of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>4.20</td>
<td>Circular genome map of Anand1_buffalo isolate of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>4.21</td>
<td>Circular genome map of Anand1_cattle isolate of <em>P. multocida</em></td>
<td></td>
</tr>
<tr>
<td>Sr. No.</td>
<td>Table No.</td>
<td>Title</td>
<td>Page No.</td>
</tr>
<tr>
<td>--------</td>
<td>-----------</td>
<td>----------------------------------------------------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>1</td>
<td>2.1</td>
<td>Diseases caused by <em>Pasteurella multocida</em></td>
<td>09</td>
</tr>
<tr>
<td>2</td>
<td>2.2</td>
<td>Gene Prediction tools, an automatic and manual annotation pipelines, databases and resources. Tools for comparative genomics/proteomics analysis</td>
<td>63</td>
</tr>
<tr>
<td>3</td>
<td>3.1</td>
<td>Details of <em>Pasteurella multocida</em> isolates from different animal species</td>
<td>67</td>
</tr>
<tr>
<td>4</td>
<td>3.2</td>
<td>Preparation of end polishing reaction mixture</td>
<td>73</td>
</tr>
<tr>
<td>5</td>
<td>3.3</td>
<td>Quality assessment of 454 library</td>
<td>77</td>
</tr>
<tr>
<td>6</td>
<td>3.4</td>
<td>Preparation of live amplification mix</td>
<td>78</td>
</tr>
<tr>
<td>7</td>
<td>3.5</td>
<td>Amount of library to be used for various molecules/bead ratio</td>
<td>79</td>
</tr>
<tr>
<td>8</td>
<td>3.6</td>
<td>Amplification cycles for small volume emPCR</td>
<td>81</td>
</tr>
<tr>
<td>9</td>
<td>3.7</td>
<td>Amplification cycles for large volume emPCR</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>3.8</td>
<td>Preparation of DNA bead incubation mixture</td>
<td>87</td>
</tr>
<tr>
<td>11</td>
<td>3.9</td>
<td>Types of beads</td>
<td>88</td>
</tr>
<tr>
<td>12</td>
<td>3.10</td>
<td>Preparation of beads for layers 1, 2 and 3</td>
<td>88</td>
</tr>
<tr>
<td>13</td>
<td>3.11</td>
<td>Packing beads and BB2 addition to DNA beads</td>
<td>89</td>
</tr>
<tr>
<td>14</td>
<td>3.12</td>
<td>Addition of titanium supplement CB</td>
<td>91</td>
</tr>
<tr>
<td>15</td>
<td>4.1</td>
<td>Stock library preparation</td>
<td>105</td>
</tr>
<tr>
<td>16</td>
<td>4.2</td>
<td>Percentage bead recovery of small volume emPCR</td>
<td>106</td>
</tr>
<tr>
<td>17</td>
<td>4.3</td>
<td>Summary of data generated by GS FLX cluster after reference mapping with Pm70</td>
<td>108</td>
</tr>
<tr>
<td>18</td>
<td>4.4</td>
<td>Summary of data generated by GS FLX cluster after <em>De</em></td>
<td>109</td>
</tr>
<tr>
<td>19</td>
<td>4.5</td>
<td>Coding sequences (CDS) in different functional groups in four isolates of <em>P. multocida</em></td>
<td>115</td>
</tr>
<tr>
<td>20</td>
<td>4.6</td>
<td>Genomic statistics of <em>P. multocida</em> isolate of different species obtained by PGAAP analysis</td>
<td>129</td>
</tr>
<tr>
<td>21</td>
<td>4.7</td>
<td>Virulence associated genes of <em>P. multocida</em> from different species of animals by PGAAP analysis</td>
<td>132</td>
</tr>
<tr>
<td>22</td>
<td>4.8</td>
<td>Unique virulence associated genes (Categorywise) in different isolates of <em>P. multocida</em></td>
<td>134</td>
</tr>
<tr>
<td>23</td>
<td>4.9</td>
<td>List of probable virulence associated genes with their function</td>
<td>168</td>
</tr>
</tbody>
</table>
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Figure No.</th>
<th>Title</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1</td>
<td>Main steps required for the whole-genome shotgun sequencing of bacterium and generation of genomic data</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>2.2</td>
<td>The pyrosequencing process</td>
<td>37</td>
</tr>
<tr>
<td>3</td>
<td>2.3</td>
<td>Whole genome shotgun sequencing</td>
<td>51</td>
</tr>
<tr>
<td>4</td>
<td>2.4</td>
<td>Workflow for high-throughput whole genome sequencing in bacteria</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>3.1</td>
<td>Workflow for the GeneMarkS+ used by NCBI database</td>
<td>99</td>
</tr>
<tr>
<td>6</td>
<td>3.2</td>
<td>Overview of gene prediction pipeline</td>
<td>99</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>%</td>
<td>Per cent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>μL</td>
<td>Micro liter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>°C</td>
<td>Degree Celsius</td>
<td></td>
<td></td>
</tr>
<tr>
<td>μg</td>
<td>Micro gram</td>
<td></td>
<td></td>
</tr>
<tr>
<td>μg/ml</td>
<td>Microgram Per Millilitre</td>
<td></td>
<td></td>
</tr>
<tr>
<td>μL</td>
<td>Micro liter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5S, 16S and 23S</td>
<td>5 Svedberg, 16 Svedberg, 23 Svedberg</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ab intio</td>
<td>From the Beginning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA</td>
<td>Blood Agar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLAST</td>
<td>Basic Local Alignment Search Tool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>Base-Pair</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDS</td>
<td>Coding Domain Sequence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPSs</td>
<td>Capsupar Polysachharide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CRISPR</td>
<td>Clustered Regularly Interspaced Short Palindromic Repeat</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DBIM</td>
<td>DNA Bead Incubation Mix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ddNTPs</td>
<td>Dideoxy Nucleotide Triphosphate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double Stranded DNA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli</td>
<td>Escherichia coli</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylene Diamine Tetraacetic Acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>emPCR</td>
<td>Emulsion Polymerase Chain Reaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>et al.</td>
<td>Et alii</td>
<td></td>
<td></td>
</tr>
<tr>
<td>etc.</td>
<td>Et cet-er-a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAO</td>
<td>Food and Agriculture Organization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIGfams</td>
<td>Fellowship of Interpretation of Genome Families</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>Gram(S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GOLD</td>
<td>Genome Online Database</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
<td></td>
<td></td>
</tr>
<tr>
<td>i.e.</td>
<td>id est (that is)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INSDSC</td>
<td>International Nucleotide Sequence Database Collaboration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IROMPs</td>
<td>Iron Regulated Outer Membrane Proteins</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lf</td>
<td>Lactoferrin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lib</td>
<td>Library</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPS</td>
<td>Lipopolysachharide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>Molar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALDI-TOF</td>
<td>Matrix Assisted Laser Desorption Ionization - Time Of Flight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MCA</td>
<td>MacConkey Agar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mg</td>
<td>Mili gram</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIDs</td>
<td>Multiplex Identifiers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>Minutes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mL</td>
<td>Mili liter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mM</td>
<td>Mili molar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mm</td>
<td>Mili meter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPC</td>
<td>Magnetic Particle Concentrator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger Ribonucleic Acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NCBI</td>
<td>National Center For Biotechnology Information</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation Sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nm</td>
<td>Nanometer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nr</td>
<td>Non Redundant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nt</td>
<td>Nucleotide</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OIE</td>
<td>Office International des Epizootics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OMP</td>
<td>Outer Membrane Protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P. multocida</td>
<td>Pasteurella multocida</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEG</td>
<td>Protein Encoding Gene</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pg</td>
<td>Pico gram</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGAAP</td>
<td>Prokaryotic Genomes Automatic Annotation Pipeline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>Hydrogen ion concentration</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM-PCR</td>
<td>Pasteurella multocida Polymerase Chain Reaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RAST</td>
<td>Rapid Annotation Subsystem Technology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL</td>
<td>Rapid Library</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rpm</td>
<td>Rotation Per Minute</td>
<td></td>
<td></td>
</tr>
<tr>
<td>sec.</td>
<td>Seconds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMRT</td>
<td>Single Molecule Real Time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP</td>
<td>Single Nucleotide Polymorphism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>spp.</td>
<td>Species</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TBE</td>
<td>Tris borate EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TF</td>
<td>Transferrin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer Ribonucleic Acid</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VF</td>
<td>Virulence Factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>viz.</td>
<td>Videlicet</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WGS</td>
<td>Whole Genome Sequencing</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### DEFINITIONS

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotation</td>
<td>The process of identifying the locations of genes and all of the coding regions in a genome and determining what those genes do</td>
</tr>
<tr>
<td>Assembly</td>
<td>It refers to the process of taking a large number of short DNA Sequences and putting them back together to create a representation of the original chromosome from which the DNA originated</td>
</tr>
<tr>
<td>Contig</td>
<td>It is a set of overlapping DNA segments that together represent a consensus region of DNA</td>
</tr>
<tr>
<td>Coverage</td>
<td>It is the number of times a nucleotide is read during the sequencing process</td>
</tr>
<tr>
<td>L50</td>
<td>It is the number of scaffolds that accounts for more than 50% of the genome assembly</td>
</tr>
<tr>
<td>Mapping</td>
<td>A method used for determining the location and relative distances between genes on a chromosome</td>
</tr>
<tr>
<td>N50</td>
<td>It is the length for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs</td>
</tr>
<tr>
<td>Read</td>
<td>Reads are obtained from the data files produced by DNA sequencing instruments. In shotgun sequencing, DNA sequences obtained from each cloned fragment of DNA and each nucleotide sequences is called a read. The reads are used later to reconstruct the original sequence</td>
</tr>
<tr>
<td>Read Length</td>
<td>It is the number of DNA bases sequenced</td>
</tr>
<tr>
<td>Scaffold</td>
<td>It consists of overlapping contigs separated by gaps of known length</td>
</tr>
<tr>
<td>Subsystem</td>
<td>Collection of functional roles jointly involved in a biological process or complex</td>
</tr>
</tbody>
</table>
INTRODUCTION
CHAPTER I
INTRODUCTION

Animal diseases are gaining incremental significance because of their potential impact on human health and economy, their importance for conservation of rare species, and their potential for insight into the co-evolution of hosts and pathogens. Due to relatively small genomes of infectious pathogens, NGSMs (Next Generation Sequencing Methods) have been applied extensively for sequencing genomes of a range of pathogens. These include viruses such as avian influenza (Hoper et al., 2009) or avian borna virus (Gancz et al., 2009), bacteria such as MRSA Staphylococcus (Highlander et al., 2007) and avian mycobacteria (Paustian et al., 2008), and protozoa such as Toxoplasma gondii (Bontell et al., 2009). These methods have also been used in genome-sequencing or genetic-association studies to identify mutations responsible for genetic disorders or diseases (ten Bosch and Grody, 2008; Vasta et al., 2009). In addition, researchers are beginning to apply NGSMs for diagnosis or identification of pathogenic organisms by random screening (Adams et al., 2009; Nakamura et al., 2008, Nakamura et al., 2009; Jones et al., 2009) and by targeted methods using PCR amplicons (Jordan et al., 2009). Targeted tagged pyrosequencing methods have also been used to assess variation in host immune-system genes, such as those of the Major histocompatibility complex (MHC) (Babik et al., 2009; Bentley et al., 2009; Wiseman et al., 2009) and immunoglobulins (Glanville et al., 2009).

Bacteria must retain their genetic information from one generation to the next. On the other hand, bacteria need to evolve strategies allowing the generation of new genetic diversity in order to survive and adapt to the continually changing environmental conditions and niches in which they live. Therefore, genetic plasticity
mirrors different bacterial lifestyles and physiological versatilities (Dobrindt and Hacker, 2001). There are three main forces which are shaping bacterial genomes, gene gain, gene loss and gene change.

Determination of entire genome sequences, however, is only the first step in understanding the inner working of an organism. The next critical step is to elucidate the functions of these sequences and give biochemical, physiological, and ecological meaning to the information. Sequence analysis indicates that the biological functions of substantial portions of complete genomes are unknown. Defining the role of each gene in the complex cellular machine and network is a formidable task (DeRisi et al., 1997; Hieter and Boguski, 1997; Strauss and Falkow, 1997). In addition, genomes contain hundreds to thousands of genes, many of which encode multiple proteins that interact and function together as multicomponent systems or apparatuses for accomplishing specific cellular processes. The products of many genes are often coregulated in complex signal transduction networks, and understanding how the genome functions as a whole to give life to complete organisms presents an even greater challenge (DeRisi et al., 1997).

The use of small subunit ribosomal RNA sequence data for understanding the evolutionary relationships of organisms was introduced by Woese in 1987. Molecular ecological studies based upon 16S rDNA sequencing, and its application to microbial ecology have revolutionised microbial ecology. 16S rDNA is too conserved to define organisms at a taxonomic level which corresponds to any appropriate concept of the species, whether a taxonomic (Goodfellow et al., 1997) or a natural (Ward, 1998) species concept. 16S rDNA sequencing has provided a phylogenetic framework to understand the bacterial diversity at the level of domains, phyla down to genera (Woese, 1987). It is the initial tool in identifying species and the only tool which
enables the identification of prokaryotic “phylotypes” and their enumeration in metagenomic studies until recently when increasing amounts of data allow identification using other techniques (Venter et al., 2004). But, 16S sequences are too conserved to effectively discriminate prokaryotes at the species level (Stackebrandt and Goebel, 1994).

The first sequence of an entire bacterial genome (Haemophilus influenzae) was published in 1995 (Fleischmann et al., 1995). Since then, the total number of publicly available genome sequences has grown rapidly (http://igweb.integratedgenomics.com/GOLD/) and there has been an increasing interest in the use of these genome sequence data to assess evolutionary relationships among bacterial species (Eisen, 2000; Wolf et al., 2002). The total number of prokaryotic genome projects running worldwide are 32,437, out of which 3,658 are completed, 767 are chromosomes with gaps, 11,034 are in scaffold and 16,978 are in contigs (http://www.ncbi.nlm.nih.gov/genome/browse/) till date. The comparison of single genes (especially the 16S rRNA gene) to infer phylogenetic relationships among bacteria has been widely used for several decades (Wheelis et al., 1992; Eisen, 1995; Garrity and Holt, 2001), but there has been considerable debate whether a tree based on any single gene can accurately represent the evolution of a species, considering the possibility of horizontal gene transfer (Lan and Reeves, 1996; Ochman et al., 2000) and the possibility of degradation of the phylogenetic signal because of saturation for amino acid substitutions (Forterre and Philippe, 1999) as complicating factors.

The study of these genomes by both computational and experimental approaches has significantly advanced understanding of the physiology and pathogenicity of many microbes and provided insights into the mechanisms and history of genome evolution (Raskin et al., 2006). The genomic DNA sequence
encodes all of the heritable information responsible for microbial replication, virulence, host specificity, and ability to evade the immune system. A comprehensive knowledge of a pathogen's genome provides all of the necessary information required for cost-effective and targeted research into disease prevention and treatment (May et al., 2001).

Whole genome amplification methods are used to generate large amounts of genomic DNA from small or precious samples. In the field of species identification, the ability to rapidly amplify genomes from limited DNA samples can significantly increase the classification accuracy by enabling the large-scale screening of several loci using high-throughput technologies. Several whole genome analysis strategies have been developed either using PCR or isothermal amplifications, each one with their own specific strengths and limitations (Pereira et al., 2008).

*Pasteurella multocida* is a small, gram-negative, non-motile, non-spore forming coccobacillus, facultative anaerobe belonging to the family *Pasteurellaceae* with bipolar staining features having 0.2-0.4 X 0.6-2.5 μm size. It is a member of the normal flora of the upper respiratory and gastrointestinal tract of many domestic and wild animals (Zurlo, 2000). *Pasteurellaceae* family includes 15 genera with four main genera namely *Pasteurella*, *Haemophilus*, *Actinobacillus* and *Lonepinella* (Hunt et al., 2000; Kuhnert and Christensen, 2008), which are obligate parasites or commensals of vertebrate animals, colonizing mainly the mucosal surface of the upper respiratory tract, oropharynx and other parts of their hosts.

*Pasteurella multocida* is a causative agent of several economically significant animal diseases occurring in numerous species viz: cattle, buffaloes, sheep, goats, poultry, turkeys, yak, deer, hippopotamus, mink, monkey, cat, elk, tortoise, snow
leopard, lion, rabbits, horses and camels (Shivachandra et al., 2011). Serious infectious diseases which include fowl cholera, bovine hemorrhagic septicemia (HS), porcine atrophic rhinitis and lapine snuffles are caused by *P. multocida*, whereas *M. haemolytica* is the causitive agent of shipping fever or pneumatic pasteurellosis. *P. multocida* strains can be differentiated into 5 serogroups (A, B, D, E, and F), based on capsule antigens, and into 16 serotypes based on lipopolysaccharide (LPS) structures. Each serogroup produces a distinct capsular polysaccharide, with serogroups A, D and F producing capsules composed of hyaluronic acid (HA) (Cifonelli et al., 1970), heparin and chondroitin (DeAngelis et al., 2002), respectively. *P. mutocida* plays a secondary role in pathogenesis of the disease or acts in combination with other bacterial, viral or parasitic infections. HS is Office International des Epizooties (OIE) listed disease and case fatality approaches 100% if treatment is not followed at the initial stage of infection in infected animals in endemic areas of Africa and Asia (De Alwis, 1999; FAO, 2005; OIE, 2009). The disease has a high global index as an animal health constraint to poor farmers (Perry et al., 2002).

The structures of the serogroup B and E capsules are not known but preliminary compositional analysis suggests that these capsules have a more complex structure than those produced by serogroups A, D and F (Muniandy et al., 1992). Fowl cholera is generally caused by the A:1, A:3 or A:4 strains (Harper et al., 2007); whereas serogroups B:2 (Asian serotype) and E:2 (African serotype) are implicated in hemorrhagic septicemia and serogroup D responsible for atrophic rhinitis (Campoy et al., 2006). Hemorrhagic septicemia is endemic in most parts of tropical Asia, Africa, and India and causes high mortality in livestock (Bain et al., 1982). Various other serotypes, namely A:1, A:1,3, A:3, A:4, B:1, B:2,5, B:3,4, E:2,5, F:3, F:3,4 have also
been isolated from HS outbreaks (De Alwis, 1999; Biswas et al., 2004; Kumar et al., 2004; Kumar et al., 2009).

Haemorrhagic septicemia is considered to be the most economically important disease of livestock in South East Asia and causes significant economic losses in India and Africa (Verma and Jaiswal, 1998). Most of the oceanic countries such as Australia, New Zealand, Japan and Western Europe have not reported the disease (De Alwis, 1999). Cattle and buffalo are the most common hosts, but pigs, sheep, goats, deer, and camels are also susceptible to infection and disease (Dawkins et al., 1991; Blackall et al., 2000). It is presumed that HS may occur in developed countries with high water buffalo population and where conditions are similar to those of tropical Asia and Africa (Benkirane and De Alwis, 2002). In India, during the past four decades, it has been found that HS accounted for 46-55% of all bovine deaths. During the 12 years period between 1974 to 1986, it accounted for 58.7% of the aggregate of deaths due to five endemic diseases, viz. foot-and-mouth disease (FMD), rinderpest, blackquarter, anthrax and HS (Dutta et al., 1990). In an active surveillance study in Sri Lanka, it was shown that in the 1970's, around 15% buffaloes and 8% cattle in the HS endemic areas died of HS annually. During the same period, the passive reporting systems recorded only 1200 to 1500 deaths a year in cattle and buffalo populations of approximately 2.5 million (De Alwis and Vipulasiri, 1981; Benkirane and De Alwis, 2002). Despite the application of advanced investigation and diagnostic techniques on the organism and the affected animal species, Pasteurella infections still continue to contribute to heavy losses in animal production.

Because of the economic relevance of P. multocida infections, research work in several fields concerning this bacterial species, such as molecular characterization of pathogenic factors and construction of mutants to be used as vaccines, is
increasing. Nevertheless, few genetic tools for *P. multocida* genetic manipulation are available (Campoy *et al.*, 2006). Little is known about the molecular basis for pathogenesis, but the genome does contain some interesting virulence functions, for example two hemagglutinins that are similar to filamentous hemagglutinins from pathogens such as *Bordetella pertussis* and *Haemophilus influenzae*. Therefore, whole genome sequencing and its analysis by the modern biotechnological and bioinformatics tools, may not only help to establish and confirm the presence and behaviour of few reported genes, but can also lead to some new insight in genome and its functional behaviour. Also, as per the literature, whole genome sequencing of *P. multocida* has apparently invited less scientific attention before, and such an scientific endeavours on local isolates of *P. multocida* from different animal species, seems to have potential for revealing novel understanding of this important pathogen causing various important diseases in animals and birds.

Considering the above perspectives, present work was undertaken with following objectives:

OBJECTIVES:

1.1 Whole genome sequence characterization and annotation of *Pasteurella multocida* from different species using pyrosequencing based 454 sequencing technology.

1.2 Identification of virulence associated genes from *Pasteurella multocida* of different species.
REVIEW OF LITERATURE
CHAPTER II
REVIEW OF LITERATURE

Bacteria are the most abundant group of organisms and a major cause of animal diseases and mortality. Bacterial cells account for most of the earth’s biomass (Whitman et al., 1998). Successful colonization of a host is essential to the life cycle of the pathogen, and the dynamics of the host-pathogen interaction determine the outcome of the interaction, including the severity of disease (Wilson, 2012).

Bacterial genomes display variation in size, even among strains of the same species. As the microorganisms have very little non-coding or repetitive DNA, the variation in genome size usually reflects differences in gene repertoire. Some species, particularly bacterial parasites and symbionts, have undergone massive genome reduction and simply contain a subset of the genes present in their ancestors (Moran, 1996). However, in free-living bacteria, such gene loss can not explain the observed disparities in genome size because ancestral genomes would have had to contain impossibly large numbers of genes. A substantial fraction of the difference in gene contents in free-living bacteria is due to the presence of ORFans, that is, open reading frames (ORFs) that have no known homologs and are consequently of no known function (Fischer and Eisenberg, 1999).

*Pasteurella multocida* is capable of causing disease in a wide range of animals and birds and subdivided into four subspecies; *multocida*, which includes the type species and three other viz. *gallicida*, *septica* and *tigris* (Boyce et al., 2010). There is a clear correlation between capsular type and the disease, with serotypes A and F typically associated with fowl cholera, serotypes B and E with haemorrhagic septicaemia in cattle and serotype D strains expressing PMT toxin with atrophic
rhinitis in swine. Unclearity is whether the expression of a particular type of lipopolysaccharide or the type and amount of proteins presented on the bacterial surface also contribute to the disease specificity (Harper et al., 2006).

Table 2.1: Diseases caused by *Pasteurella multocida*

<table>
<thead>
<tr>
<th>Disease</th>
<th>Capsular serogroup</th>
<th>Species affected</th>
<th>Virulence factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animal bite associated sepsis</td>
<td>Usually not determined,</td>
<td>Any, but most commonly humans, cats, dogs</td>
<td>Not established</td>
</tr>
<tr>
<td></td>
<td>but where given usually</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A, occasionally D</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemorrhagic septicaemia</td>
<td>B and E</td>
<td>Cattle, buffalo, pigs, goats, deer,</td>
<td>Capsule, fimbriae, filamentous haemagglutinin</td>
</tr>
<tr>
<td></td>
<td></td>
<td>camels, rarely horses</td>
<td></td>
</tr>
<tr>
<td>Avian cholera</td>
<td>A, F and very rarely D</td>
<td>Chickens, turkeys, ducks, wild birds</td>
<td>Capsule, lipopolysaccharide (LPS), iron acquisition proteins, fimbriae, filamentous haemagglutinin, sialic acid uptake</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atrophic rhinitis</td>
<td>D and rarely A</td>
<td>Pigs and rabbits</td>
<td><em>Pasteurella multocida</em> toxin (PMT)</td>
</tr>
<tr>
<td>Snuffles</td>
<td>A and D</td>
<td>Rabbits</td>
<td>Not established</td>
</tr>
<tr>
<td>Enzootic pneumonia, shipping fever</td>
<td>A, D and rarely F</td>
<td>Cattle, sheep and pigs</td>
<td>Not established. Co-infection with other bacterial or viral species important</td>
</tr>
</tbody>
</table>

The cell surface protects Gram-negative bacteria against a range of harsh environments and is critical for interaction of the bacterium with the host. Virulent *P. multocida* strains produce a polysaccharide capsule. The serogroups A and B capsule contains hyaluronic acid, while serogroups D and F capsules contain heparin and chondroitin, respectively. The outer membrane (OM) functions as a selective barrier that prevents the entry of many toxic molecules into the cell, a property that is crucial for bacterial survival in many environments (Hatfaludi et al., 2010).
P. multocida is a commensal organism in many mammalian hosts. However, it can potentially cause disease in almost all species if it gains access to the bloodstream. Pathogenesis is a result of complex interactions between specific host factors (e.g. species, age, immune status) and specific bacterial virulence factors (e.g. LPS, capsule, adhesins, OM proteins). Therefore, disease pathogenesis is versatile and depends on the bacterial strain, the animal model and the changing response of both the host and bacteria to the interaction (Boyce et al., 2010).

Little progress is made in understanding exactly how P. multocida invades mucosal surfaces to gain access to the blood and how the host responds to the infection. However, significant advances have been made in identifying bacterial factors critical to P. multocida pathogenesis. The capsule and lipopolysaccharide are essential for normal disease progression in fowl cholera and there is some progress in understanding the bacterial response to the in vivo environment (Boyce et al., 2004) and identifying bacterial factors required for disease progression (Fuller et al., 2000; Harper et al., 2003).

2.1 VIRULENCE FACTORS OF P. multocida

Discovering virulence factors of pathogenic bacteria is a key in understanding pathogenesis and for identification of targets for novel drugs and design of new vaccines. Comparative genomics, transcriptomics and proteomics have become the popular tools in discovering the virulence factors in bacterial pathogens. Once the putative virulence factors are identified by genomics and/or proteomics, their functions and mechanisms can be further investigated by phenotypic analyses including mutagenesis, biochemical methods and/or structural biology. Combination
of these techniques will accelerate the developments of therapeutic drugs and vaccines in combating bacterial diseases (Wu et al., 2008).

Comparative genomics is a popular tool to identify virulence factors and genes involved in environmental persistence of pathogens. The goal is to correlate those differences to biological function and to gain insight into selective evolutionary pressures and patterns of gene transfer or loss, particularly within the context of virulence in pathogenic species. Comparisons can be performed either with genome sequence or by using microarray-based methods (Raskin et al., 2006).

In essence, the ability of pathogenic bacteria to cause disease in a susceptible host is determined by multiple virulence factors acting individually or together at different stages of infection. Virulence factors are often involved in direct interactions with the host tissues or in concealing the bacterial surface from the host’s defense mechanisms. Chen et al. (2005) created a database called virulence factor database (VFDB) (http://www.mgc.ac.cn/VFs/). Bacterial genome sequences rapidly add candidate virulence genes to such databases.

Bacterial virulence factors can be divided into several groups on the basis of the mechanism of virulence and function. These are (1) membrane proteins, which play roles in adhesion, colonization and invasions, promote adherence to host cell surfaces, are responsible for resistance to antibiotics, and promote intercellular communication; (2) Polysaccharide capsules that surround many bacterial species and have anti-phagocytic property; (3) Secretory proteins, such as toxin, which can modify the host cell environment and are responsible for some host cell-bacteria interactions (Wu et al., 2008).
There are many virulence factors which are responsible for the diseases due to *Pasteurella multocida* in animals and birds, which are described below.

### 2.1.1 Capsule

Capsules are a common bacterial virulence determinant and are associated with pathogenesis and immunity in many genera of bacteria (Finlay and Falkow, 1989; Moxon and Kroll, 1990). *Pasteurella multocida* can be classified by serological methods into five capsule groups designated A, B, D, E and F. The composition and structure of the capsular material found in *P. multocida* serotypes A, D and F are very similar to mammalian glycosaminoglycans and consist mainly of hyaluronan, heparosan and unsulphated chondroitin, respectively (Pandit and Smith, 1993; Rimler, 1994; DeAngelis, 1996; DeAngelis and Padgett-McCue, 2000). The genes required for the synthesis and transport of these capsular types are encoded within a single region on the genome (Townsend *et al.*, 2001). In type A, D and F strains, an additional gene encoding a heparosan synthase was identified outside of the known capsule biosynthesis region. The synthase encoded by this gene, renamed *hssB* (formerly *pglA*), is transcribed 10-fold less than the synthase within the capsule operon, uses a different acceptor and gives rise to smaller molecular weight polymer products and the expression of this gene may give rise to capsular variation (DeAngelis and White, 2004).

Generally strains that possess a capsule are more virulent than their acapsular variants (Heddleston *et al.*, 1964; Snipes *et al.*, 1987; Tsuji and Matsumoto, 1989). The important role of capsule in the pathogenesis of *P. multocida* has been clearly demonstrated as genetically defined, acapsular mutants constructed from both serogroup A and B strains were strongly attenuated in mice (Boyce and Adler, 2000;
Chung et al., 2001). The serogroup A acapsular mutant was also shown to be avirulent in chickens and was unable to establish growth in chicken muscle (Chung et al., 2001). Despite its apparent lack of persistence, vaccination of chickens with high doses of this acapsular mutant stimulated protective immunity (Chung et al., 2005). Immunity is generally serotype specific. Empirically derived, live, avirulent strains have been used as vaccines in both poultry and cattle (Derieux, 1984; Myint et al., 1987), but the basis for attenuation is not known and therefore reversion to virulence occurs (Hofacre and Glisson, 1986). However, a major advantage of live, attenuated vaccines is the fact that they are able to stimulate cross-protection against different *P. multocida* serotypes.

Capsule plays a significant role in resistance to phagocytosis, and this has been demonstrated *in vitro* by Harmon et al. (1991) and others, who have correlated sensitivity to phagocytosis with the presence and thickness of the bacterial capsule (Truscott and Hirsh, 1988; Harmon *et al*., 1991; Pruimboom *et al*., 1996). Studies using murine macrophages clearly demonstrated that a genetically defined acapsular serotype B mutant was more susceptible to uptake than its wild-type parent (Boyce and Adler, 2000). Resistance to complement-mediated lysis is clearly important for virulence and experiments on *P. multocida* type A strains have shown that serum resistance correlates with the possession of a capsule (Snipes and Hirsh, 1986; Hansen and Hirsh, 1989). A genetically defined acapsular mutant was no longer serum resistant in normal avian serum compared with the serotype A wild-type parent and complemented mutant (Chung *et al*., 2001). Jaques *et al.* (1993) reported non-capsulated variants of atrophic rhinitis strains of *P. multocida*, which were shown to have reduced virulence for mice and piglets.
Arunugam et al. (2011) isolated 114 strains of Pasteurella multocida from different domestic animal species (cattle, buffalo, sheep, goat, pig, rabbit, dog, cat), avian species (chicken, duck, turkey) and wild animals (deer, tiger, orang utan, marmoset). The serogroups of *P. multocida* were determined by both conventional capsular serotyping and a multiplex PCR assay targeting specific capsular genes. Based on the conventional serotyping method, 114 strains of *P. multocida* were subtyped into 55 species-specific (untypeable strains) *P. multocida*, 15 serogroup A, 23 serogroup B and 21 serogroup D. Based on the multiplex PCR assay on the specific capsular genes associated with each serogroup, 114 strains were further divided to 22 species-specific *P. multocida* (KMT1 - 460 bp), 53 serogroup A (A - 1,044 bp), 33 serogroup B (B - 760 bp) and 6 serogroup D (D - 657 bp). No serogroup E (511 bp) or F (851 bp) was detected among the Malaysian *P. multocida* isolates.

### 2.1.2 Lipopolysaccharide

Lipopolysaccharide stimulates humoral immunity and is considered to be a protective antigen. Wijewardana et al. (1990) raised monoclonal antibodies (MAb) against the lipopolysaccharide from a serotype A strain, which were bactericidal and protected mice against homologous challenge. An opsonic monoclonal antibody against lipopolysaccharide from a serotype B strain of *P. multocida* partially protected mice against *P. multocida* infection (Ramdani and Adler, 1991). Intravenous inoculation of lipopolysaccharide from serotype B:2 strains could reproduce clinical signs of haemorrhagic septicaemia in buffalo (Horadagoda et al., 2002), but the endotoxic properties of lipopolysaccharide from serotype A strains was less clear. Various studies have indicated that chicken embryos and mice are highly susceptible, but that turkey poults are relatively resistant (Ganfield et al., 1976; Rhoades and Rimler, 1987; Mendes et al., 1994).
P. multocida requires a complete lipopolysaccharide structure in order to replicate in vivo and cause disease (Harper et al., 2004). A galE mutant of P. multocida had significantly reduced viability in mice. The role of galE in bacteria is the epimerization of UDP-glucose to UDP-galactose before lipopolysaccharide assembly, and this mutant probably expressed an altered lipopolysaccharide (Fernandez de Henestrosa et al., 1997).

The complete lipopolysaccharide structure of strains VP161 and X73 belonging to Heddleston serotype 1 and Pm70, a Heddleston serotype 3 strain, was determined and all possessed a structure similar to the lipopolysaccharide or lipo-oligosaccharide (LOS) of Neisseria spp. and Haemophilus spp. The P. multocida lipopolysaccharide isolated from the two Heddleston type I strains contains terminal phosphocholine residues, which in other bacteria plays a key role in bacterial adhesion to, and invasion of, host cells by binding directly to the platelet-activating factor (PAF) receptor (Cundell et al., 1995; Schenkein et al., 2000; Swords et al., 2000; Serino and Virji, 2002).

In addition to phosphocholine residues, all the P. multocida strains have phosphoethanolamine residues attached to various sites on their lipopolysaccharide (St Michael et al., 2005). In other Gram-negative bacteria, phosphoethanolamine is added to a number of sites within the inner core of lipopolysaccharide by specific transferases, and in Neisseria meningitidis, the expression and position of these residues affects the ability of the bacteria to resist the innate immune response (Ram et al., 2003).
2.1.3 Fimbriae and adhesins

Attachment to host cells and/or extracellular matrix proteins is a primary prerequisite for bacterial infections; adhesins that mediate such adherence are potential virulence factors. There are many genes, including ptfA, fimA, flp1, flp2, hsf-1 and hsf-2 on P. multocida genome that encode proteins similar to fimbriae or fibrils in other bacteria. Fimbriae play a role in surface adhesion, as fimbriae have been observed on some P. multocida serotype A strains that were able to adhere to mucosal epithelium, but not on the surface of those strains unable to adhere (Glorioso et al., 1982; Rebers et al., 1988; Isaacson and Trigo, 1995; Ruffolo et al., 1997). Type IV fimbriae (pili) have been isolated and characterized from P. multocida serotypes A, B and D (Ruffolo et al., 1997) and are often associated with virulence in other bacteria because of their role in attachment to host cell surfaces. Type IV fimbriae have been used as vaccines against ovine footrot and bovine keratoconjunctivitis caused by Dichelobacter nodosus and Moraxella bovis, respectively (Hatfaludi et al., 2010).

The subunit gene, ptfA, has been isolated and sequenced from a number of strains and the predicted protein sequences showed significant variation between strains (Doughty et al., 2000). Siju et al. (2007) cloned, sequenced and expressed in E. coli the ptfA gene from P. multocida serogroup B:2. Sequence alignment showed that the P. multocida B:2 and A:1 genes were 78.4% identical at the nucleotide level, with the first 200 bp being 100% identical. Rabbit antiserum raised against the recombinant PtfA (serotype B:2) protein detected an 18 kDa protein in P. multocida B:2, A:1 and Pseudomonas aeruginosa whole cell lysates. However, the role of fimbrial structures in P. multocida virulence is still unproven.
P. multocida Pm70 genome contains a region encoding proteins with significant similarity to the Flp pilin locus in Actinobacillus actinomycetemcomitans, encoding a type IV fimbrial subfamily (Kachlany et al., 2001a; May et al., 2001). This locus encodes proteins predicted to be the Flp pilin subunits and the proteins required for the pilin assembly. Flp pili gene products in P. multocida are required for virulence in mice. Two genes, flpl, encoding a Flp pilin subunit and tadD, predicted to encode a component of the secretion apparatus required for Flp pilin assembly, have been inactivated in two independent signature tagged mutagenesis (STM) studies and both mutants were significantly attenuated in mice (Kachlany et al., 2001a; Kachlany et al., 2001b; Harper et al., 2003; Wilkie et al., 2012).

Two P. multocida genes, pJhaB1 and pfhaB2, share significant similarity with a class of genes that encode filamentous haemagglutinins, which in Bordetella pertussis play a major role in the colonization of the upper respiratory tract (Kimura et al., 1990; Mooi et al., 1992). Mutation of these genes in P. multocida resulted in significantly reduced virulence in mice (Fuller et al., 2000). A pfhaB2 mutant was constructed in a fowl cholera strain, P1059, and shown to be highly attenuated in turkeys when administered intranasally, but only moderately attenuated when given intravenously, suggested a significant role of pfhaB2 in initial colonization or invasion (Tatum et al., 2005).

2.1.4 Toxins

Most P. multocida strains that cause fowl cholera, haemorrhagic septicaemia or pneumonia are not known to express any toxins. The dermonecrotic toxin, PMT, expressed mainly by serogroup D strains, is the only toxin identified to date and is responsible for the clinical and pathological signs of atrophic rhinitis (Foged et al.,
1987; Rimler and Rhoades, 1989). Both purified native and recombinant PMT toxin can be used to experimentally induce clinical signs of the disease (Foged et al., 1987; Lax and Chanter, 1990).

In pigs, this toxin leads to atrophy of nasal turbinates, where bone resorption occurs due to uncontrolled proliferation of osteoclasts, and regeneration of bone is prevented by the inhibition of osteoblasts (Sterner-Kock et al., 1995; Mullan and Lax, 1998). PMT is also a potent mitogen, inducing many cellular effects including rearrangements in the actin cytoskeleton (Zywietz et al., 2001). Like cholera and pertussis toxins, PMT activates dendritic cells to mature cells but, unlike the other toxins, it is a poor adjuvant and appears to suppress the antibody response (Bagley et al., 2005). PMT blocks chemotaxis-induced migration of dendritic cells to regional lymph nodes and might, therefore, in a natural infection, limit the development of an adaptive immune response (Blocker et al., 2006).

The PMT toxin gene resides on a lysogenic bacteriophage belonging to the Siphoviridae family. As the PMT toxin has no signal sequence and no known mechanism of export, the lytic phase of the bacteriophage may mediate release of the toxin (Pullinger et al., 2004). The toxin is an effective immunogen, a toxoid developed by deletion mutagenesis of the cloned PMT toxin gene was found to protect mice and their offspring against challenge with purified PMT (Petersen et al., 1991). A genetically modified PMT toxin, where there were two key amino acid substitutions, led to a non-toxigenic protein that protected pigs against experimental challenge with the wild-type strain (To et al., 2005).
2.1.5 Iron regulated and iron acquisition proteins

Iron is an essential element which must be acquired by bacteria in order to survive. Because of its inherent toxicity, the level of free iron available in vivo is very limited and *P. multocida*, like other bacterial species, has developed multiple mechanisms for iron uptake. Sequence analysis of *P. multocida* Pm70 revealed that a relatively large proportion of the genome (over 2.5%) encodes 53 proteins with similarity to proteins involved in iron uptake or acquisition (May *et al.*, 2001). Whole genome microarray studies have identified a large number of genes (up to 12%) with altered expression under iron-limiting conditions or in response to defined sole iron sources (Paustian *et al.*, 2001; Paustian *et al.*, 2002). Analysis of the outer membrane sub-proteome under iron-limiting conditions identified only two proteins, OmpW and Pm803, with increased expression (Boyce and Adler, 2006).

Iron, essential for bacterial growth, is largely unavailable in vivo as a result of binding by host proteins such as transferrin and lactoferrin. Pathogenic bacteria, such as *P. multocida*, overcome these severe iron limitations by expressing a number of iron regulated proteins and low molecular weight siderophores which are able to sequester iron from the host proteins. Proteins which are expressed under iron-limiting conditions may include outer membrane proteins (OMPs) involved in iron acquisition or other virulence factors such as toxins and haemolysins (Chawak *et al.*, 2001).

Comparisons of *P. multocida* grown in iron-rich, iron-depleted media or in vivo has demonstrated that many high molecular weight outer membrane proteins are regulated by iron levels and have therefore been called iron-regulated outer membrane proteins (IROMPs) (Snipes *et al.*, 1988; Choi-Kim *et al.*, 1991). *P. multocida* grown
under iron-limited conditions induces a stronger protective response in mice compared with the same strain grown under iron-replete conditions (Kennett et al., 1993). IROMPs therefore play a significant role in cross-protective immunity (Glisson et al., 1993; Ruffolo et al., 1998).

Transferrin receptors utilized by bacterial species in Pasteurellaceae and Neisseriaceae families usually consist of two iron binding receptors TbpA and TbpB (Gray-Owen and Schryvers, 1996). The transferrin receptor in bovine strains of *P. multocida* is composed of only a single protein TbpA (Ogunnariwo and Schryvers, 2001). However, this receptor may not be present in all *P. multocida* strains. PCR and DNA hybridization study found that the *tbpA* gene is present only in some bovine and ovine clinical isolates (Ewers et al., 2006).

Iron acquisition proteins are believed to play a role in the disease process; serogroup A mutants with inactivated ExbB, ExbD, TonB or HgbA proteins are attenuated in mice (Fuller et al., 2000; Bosch et al., 2002b). ExbB, ExbD and TonB are part of the TonB transport complex, required to transport and provide energy for iron sequestration and HgbA is predicted to be a haemoglobin-binding protein required for the acquisition of iron from host proteins (Bosch et al., 2002b). Expression of *P. multocida* *hgbA* gene increased six fold under iron limitation (Paustian et al., 2001) and also during growth in blood of infected chickens (Boyce et al., 2002). Several of these iron-acquisition proteins have been tested as potential vaccine antigens; none has elicited protection (Boyce et al., 2010).

Haemoglobin-binding protein of 113.5 kDa, designated HgbB from *P. multocida* was identified by Cox et al. (2003). Bioinformatics analysis of HgbB revealed the presence of putative Fur and TonB binding sites. However, inactivation
of the hgbB did not result in reduced bacterial virulence in mice, nor did it decrease the ability of P. multocida to bind haemoglobin. Furthermore, when used as a vaccine antigen, recombinant insoluble HgbB did not confer any protection to mice against wild-type infection.

The iron component of haemoglobin, haem, can be utilised as a source of essential iron by many microorganisms. Pathogenic bacteria commonly secrete exotoxins such as haemolysin that can lyse cells and release haem. P. multocida has been shown to lyse erythrocytes under anaerobic conditions (Krewulak and Vogel, 2008).

Adler et al. (1999) demonstrated that OMPs produced under iron-limiting conditions, mimic the concentration of iron in vivo, could induce cross-protective immunity in mice whereas, OMPs produced under iron replete conditions stimulated only homologous protection (Ruffolo et al., 1998).

2.1.6 Sialic acid metabolism

Many pathogenic bacteria have evolved mechanisms to evade recognition by the immune system based on the surface exposure of sialic acid. A common theme in this host-mimicry is the incorporation of sialic acid as a component of either the capsular polysaccharide or LPS. It has been hypothesised that sialidases contribute to the virulence of some pathogenic organisms, especially those that inhabit and invade mucosal surfaces (Schauer, 2009). Sialidases act to remove sialic acid from host glycosylated proteins and lipids for use as a carbon source. These enzymes also enhance bacterial virulence by unmasking key host receptors and/or reducing the effectiveness of host defences such as mucin. Most P. multocida strains produce
sialidase both as cell bound and as extracellular enzymes (Scharmann et al., 1970; Drzeniek et al., 1972; White et al., 1995).

Two sialidases NanH and NanB have been cloned and characterized from a fowl cholera isolate of *P. multocida* (Mizan et al., 2000). These sialidases differed in their specificity, with both able to utilize 2,3′ sialyl lactose, but only NanB was able to fully utilize 2,6′ sialyl lactose. The presence of two sialidases with slightly different specificities would enhance the metabolic capacity of *P. multocida* in the host (Mizan et al., 2000).

*P. multocida* is capable of scavenging sialic acid from the environment for both the sialylation of cell components and for nutrients via a catabolic breakdown pathway. The uptake, but not catabolism, of sialic acid was shown to be essential for virulence in mice (Steenbergen et al., 2005). Two genes, *pm0188* and *pm0508*, encoding sialic acid transferases, have been identified in the *P. multocida* genome (Yu et al., 2005).

### 2.1.7 Hyaluronidase

Although, the role of hyaluronidase in pathogenesis of *P. multocida* infections has not been determined, it is present in many of the serotype B strains of *P. multocida* that cause bovine haemorrhagic septicaemia.

A study of 74 *P. multocida* strains representing all capsular serotypes found that only the type B strains, isolated from haemorrhagic septicaemia infections, produced hyaluronidase (Carter and Chengappa, 1980). Another study of 176 strains of *P. multocida* representing different serotypes also found hyaluronidase activity confined to serotype B, but more specifically B:2, and it was suggested that a test for
2.1.8 Outer membrane proteins

Outer membrane proteins (OMPs) have been separated into OM integral and lipoproteins, and then further separation of these two groups into functional categories. Bacterial OM lipoproteins have been grouped according to functional type either as structural proteins, enzymes, receptors or transporters performing essential functions at the membrane-aqueous interface (Babu et al., 2006). For integral membrane proteins, the functional categories include: small b-barrel membrane anchors, nonspecific porins, specific channels, energy-dependent transport-efflux, energy-dependent transporter-influx, protein secretion pore, OM usher proteins, adhesins, membrane integral enzymes and protein pore-forming toxins (Nikaido, 2003; Wimley, 2003).

A protein of 39 kDa was identified in *P. multocida* A:3 strain P1059 and its expression was shown to correlate with the presence and amount of capsule present on the cell (Borrathybay et al., 2003; Ali et al., 2004). *P. multocida* adhere and invade chicken embryo fibroblasts, and this adherence was inhibited by both monoclonal and polyclonal antibodies raised against the 39 kDa protein (Borrathybay et al., 2003; Ali et al., 2004; Ali et al., 2004). Passive immunization of mice with a monoclonal antibody against the 39 kDa protein or active immunization with affinity purified 39 kDa protein, demonstrated that antibodies raised against this protein were cross-protective against serovars A:1 and A:3 (Ali et al., 2004; Ali et al., 2004). This protein was identified as PlpB (*Pasteurella* lipoprotein B), using peptide mass fingerprinting (Tabatabai and Zehr, 2004) and is predicted to be an ABC (ATP-
binding cassette) transport protein required for the uptake of methionine into the cell (Merlin et al., 2002).

One of the major outer membrane proteins of *P. multocida* is OmpH. Antibodies raised against this protein provide some protection against the disease. Monoclonal antibodies specific for OmpH passively protected mice against *P. multocida* challenge (Marandi and Mittal, 1997) and vaccination with the native, but not recombinant, OmpH protein elicited protective immunity in birds against homologous challenge (Luo et al., 1997). In addition, antibodies raised to an OmpH synthetic peptide, Cyclic-L2, provided partial protection in chickens against homologous challenge (Luo et al., 1999). OmpH expression is regulated by the Fur (ferric uptake regulator) protein and responds to glucose and iron concentrations (Bosch et al., 2001). Garrido et al. (2008) identified two copies of OmpH (OmpH1 and OmpH2) in serogroup A strain PM108.

Examining the ability of *P. multocida* to bind host extracellular matrix proteins have shown that the bacteria can adhere to fibronectin and collagen type IX. Proteins identified as possible adhesins include OmpA, Oma87, Pm1069 and the iron related proteins, Tbp (transferrin binding protein) and the putative TonB receptor HgbA (Dabo et al., 2005). Bovine *P. multocida* strains are able to utilise iron from bovine transferrin, but not from transferrin of other species, whereas, avian strains are unable to capture iron from any transferrin molecules including those from avian source (Hatfaludi et al., 2010). *P. multocida* has a single, novel TbpA receptor capable of efficiently mediating iron acquisition from bovine transferrin without the involvement of a second receptor protein TbpB.
Veken et al. (1994) examined *P. multocida* B:2,5 haemorrhagic septicaemia strains for TbpA homologues. An 82 kDa protein was present in B:2,5 strains and was found to bind bovine transferrin. However, no proteins that bound transferrin could be isolated from a B:3,4 serotype strains. Shivachandra et al. (2005) amplified the *tbpA* gene from chicken (A:1), buffalo (B:2), pig (D:1) and duck (F:3) *P. multocida* strains. Sequence analysis of the *tbpA* gene from serotype B:2 showed 98% identity to the A:1 strain. However, Ewers et al. (2006) analysed 289 strains for distribution of virulence genes and no *tbpA*-specific PCR fragment was amplified from avian or pig strains. The *tbpA* gene was present in only 31.5% of the strains. In this study, *tbpA* was found exclusively in isolates from bovine (70.2% of all tested bovine strains), sheep (80%) or buffalo (57.1%) strains.

Another 30.1 kDa PlpB protein involved in methionine binding, represents the periplasmic binding component of a methionine uptake ABC transporter. PlpB was initially identified as a cross-protective antigen, recombinant PlpB failed to protect either mice or chickens against *P. multocida* challenge, despite stimulating an antibody response (Wu et al., 2007), but later found that the cross-protective antigen is PlpE. Immunisation with recombinant *P. multocida* PlpE conferred protective immunity, with 80-100% of mice and 63-100% of chickens protected against heterologous challenge (Wu et al., 2007). PlpE is the first *P. multocida* recombinant protein to stimulate high level cross-serotype protective immunity.

2.1.9. Transcriptional regulators

Fis is a growth phase-dependent, nucleoid-associated protein which plays a role in the transcriptional regulation of a number of genes in diverse bacterial species (Grainger and Busby, 2008). In *Escherichia coli*, Fis is expressed at high levels in...
actively growing cells (50,000 molecules per cell in early exponential growth phase) and expression drops to very low levels during stationary phase (Ball et al., 1992; Bradley et al., 2007). In addition to growth phase regulation, levels of Fis are negatively regulated by the stringent response during nutrient starvation (Ninnemann et al., 1992). Fis can act as both, a positive or negative regulator of transcription and it has both direct and indirect effects on gene transcription. In E. coli and Salmonella, Fis binds to a degenerate 15-bp consensus sequence GNtYAAWWaTRaNC, inducing DNA bending, but only a few of the sequences fitting this consensus are high affinity binding sites (Pan et al., 1996; Shao et al., 2008). Fis is involved in the regulation of genes encoding a wide range of functions, including quorum sensing in Vibrio cholerae (Lenz and Bassler, 2007), and certain virulence factors in pathogenic E. coli (Goldberg et al., 2001; Kelly et al., 2004) and Salmonella (Kelly et al., 2004) and Erwinia chrysanthemi (Saldana et al., 2009). The absence of functional Fis protein led to the loss of capsule expression in P. multocida. The presence of functional Fis protein is also required for the expressions of virulence gene pfhB2 and its predicted secretion partner lspB2 in P. multocida (Steen et al., 2010).

Other virulence factors such as neuraminidases, iron sequestering proteins and metabolic enzymes play key roles in acquiring and utilizing substrates for growth within the host, often a relatively nutrient poor and hostile environment (Harper et al., 2006).

Adler et al. (1999) identified potential protective antigens or virulence genes to be used as candidates for attenuating mutations or as the basis for vaccine antigen delivery systems. The gene encoding an outer membrane protein, Oma87, Type 4 fimbriae, fimbrial subunit protein PtfA, four bex genes involved in capsule transport, transferrin binding protein Tbp1, mesA gene responsible for haemolytic phenotype
under anaerobic conditions were studied using different approaches for their use as candidate vaccine antigens.

Ewers et al. (2006) isolated 289 strains from *Pasteurella multocida* wild type and reference strains viz. cattle (104), small ruminants (15), buffaloes (7), swine (52), rabbits (20), poultry (20), cats (54), dogs (8), human (2), and of unknown host origin (7) from various clinically healthy and diseased hosts for detecting capsule biosynthesis genes (*capA, B, D, E and F*) and 14 virulence associated genes. Capsule type A strains were highly adapted to bovines (92.3%) and poultry (85.7%), while *capA* (34.9%) and *capD* (58.1%) positive strains in swine. *capD* positive strains also from small ruminants (34.9%) and *capF* was detected in wild type strains from diseased cattle (2.2%) and cats (7.4%). None of the isolates harboured *capE*, while *capB* was exclusively found in all strains from buffaloes. Nearly all isolates showed a combination of genes encoding outer membrane proteins, colonization factors, iron acquisition factors and superoxide-dismutases without any clue for host specificity. In contrast, the transferrin binding protein encoding gene *tbpA* (31.5%) was limited to ruminant strains and only 37.0% of all *P. multocida* strains harboured *pfhA*, coding for a filamentous hemagglutinin, supposed to be a putative adhesion and serum resistance factor. *PfhA* revealed a strong positive association to the outcome of disease in bovine hosts and in combination with *toxA* to that in swine. The dermonecrototoxin encoding *toxA*, present in 12.5% of all strains, was detected in isolates from swine, small ruminants, cattle and poultry. A significant association to the disease status, however, was only existent in swine, although with 66.7% found a notably high prevalence of the toxin gene among strains from small ruminants. The genes *toxA, bpA* and *pfhA* as well as capsule biosynthesis genes were supposed to be important epidemiological marker genes for characterizing *P. multocida* field strains.
Atashpaz et al. (2009) developed a reliable multiplex PCR method for rapid
detection of four virulence genes, i.e., *tbpA*, *pfliA*, *toxA* and *hgbB* to determine the
epidemiological correlation between *P. multocida* infection and existence of virulence
genes in *P. multocida*. Eighty seven strains of *P. multocida* isolated from various
clinically healthy and infected hosts were examined by uniplex PCR method for each
virulence associated genes. For rapid detection of four virulence genes, they
developed a multiplex PCR method to benefit the epidemiological investigations.

Tang et al. (2009) isolated a total of 233 of *P. multocida* isolates obtained
from 2,912 cases of clinical respiratory disease in pigs in China. Serogroup A, *P.
multocida* isolates were obtained from 92 cases (39.5%), and serogroup D isolates
from 128 cases (54.9%); 12 isolates (5.2%) were untypeable. All the isolates were
characterized for their susceptibilities to 20 antibiotics and the presence of 19 genes
for virulence factors (VFs). Use of PCR showed that colonization factors (*ptfA*, *fimA*,
and *hsf-2*), iron acquisition factors, sialidases (*nanH*) and outer membrane proteins
occurred in most porcine strains. The VFs *pfhA*, *tadD*, *toxA*, and *pmHAS* were each
present in <50% of strains. Various VFs exhibited distinctive associations with
serogroups: concentrated in serogroup A, concentrated in serogroup D, or occurring
jointly in serogroups A and D.

Garcia et al. (2011) obtained 205 *P. multocida* isolates from pigs, which were
phenotypically and genetically characterised by determining their biovar, capsular
type, virulence-associated genes and pulsed-field gel electrophoresis (PFGE) profiles.
All the isolates were identified as *P. multocida* subspecies *multocida* and most were
assigned to biovar 3 (58%) and biovar 2 (39.5%). Biovar 1 represented 24% of the
isolates. According to the capsular type, a great majority of the isolates (79.0%)
belonged to capsular type A, 18.5% belonged to capsular type D and 24% were of
capsular type F. All the isolates harboured *ompH*, *psl*, *oma87*, *ptfA*, *nanB*, *nanH*, *tonB*, *hgbA*, *sodA* and *sodC* genes, while none of them possessed the transferrin-binding protein gene *tbpA*. The prevalence of *toxA*, *pfhaA* and *hgbB* genes was variable (7.8, 40.5 and 60.5% of the isolates, respectively). After PFGE typing, isolates of biovar 2 and 3 were grouped in two different clusters (A and B) at a level of 45 per cent similarity. In addition, isolates of biovar 2 and 3 exhibited statistically significant differences (*P*<0.05) in the virulence-associated *hgbB* and *pfhaA* genes (biovar 3 was *hgbB*+ *pfhaA*-, while biovar 2 was *hgbB*-*pfhaA*+).

**Furian et al.** (2013) investigated 12 genes related to virulence in 25 samples of *P. multocida* isolated from fowl cholera cases in southern region of Brazil through the development of multiplex PCR protocols. The *ompH*, *oma87*, *sodC*, *hgbA*, *hgbB*, *exBD-tonB* and *nanB* genes were present in 100% of the samples (25/25), the *sodA* and *nanH* genes were present in 96% (24/25), *ptfA* was present in 92% (23/25), and *pfhaA* was present in 60% (15/25). Gene *toxA* was not identified in any of the samples. Five different genetic profiles were obtained, of which PI (negative to *toxA*) was the most common.

**Verma et al.** (2013) performed isolation of 23 isolates of *P. multocida*, obtained from 335 cases of various clinically healthy and diseased cattle. These isolates were examined for capsule synthesis genes (*capA*, *B*, *D*, *E* and *F*) and 11 virulence associated genes (*tbpA*, *pfhaA*, *toxA*, *hgbB*, *hgbA*, *nanH*, *nanB*, *sodA*, *sodC*, *oma87* and *ptfA*) by PCR. A total of 19 *P. multocida* isolates belonging to capsular type B and 4 of capsular type A were isolated. All the isolates of capsular type B harboured the virulence associated genes: *tbpA*, *pfhaA*, *hgbA*, *sodC* and *nanH*, coding for transferrin binding protein, filamentous hemagglutinin, haemoglobin binding protein, superoxide dismutase and neuraminidases, respectively; while isolates
belonging to capsular type A also carried *tbpA*, *pfhA*, *hgbA* and *nanH* genes. Only 50% of capsular type A isolates contained *sodC* gene while 100% of capsular type B isolates had *sodC* gene. Genes *nanB* and *toxA* were absent in all the 23 isolates. In capsular type A isolates, either *sodA* or *sodC* gene was present and these genes did not occur concurrently. The presence of virulence associated gene *ptfA* revealed a positive association with the disease outcome in cattle and could therefore be an important epidemiological marker gene for characterizing *P. multocida* isolates.

Jamali *et al.* (2014) isolated *Pasteurella multocida* from calves with respiratory infection in Iran. *P. multocida* was detected in 141/169 bovine respiratory infection cases on Iranian dairy and beef farms. *P. multocida* were grouped into serogroups A (126/141), D (12/141), and B (3/141). Of the *P. multocida* isolates, all harboured *psl, ompH, oma87, fimA, ptfA, nanB* and *nanH* genes, 139/141 had *hsf-2*, and 115/141 had *pfhA*, and *tadD*.

Khamesipour *et al.* (2014) isolated 30 *Pasteurella multocida* strains from 333 pneumonic and apparently health slaughter cattle, examined for capsule biosynthesis genes and 23 virulence-associated genes by polymerase chain reaction (PCR). Of the isolates, 23 belonged to capsular type A, 5 to capsular type D and 2 isolates were untypeable. The distribution of the capsular types in pneumonic lungs and in apparently healthy lungs was statistically similar. All the virulence genes tested were detected among the isolates derived from pneumonic lungs; whereas isolates derived from apparently healthy lungs carried 16 of the 23 genes. The frequently detected genes among the isolates from pneumonic lungs were *exbD, hgbA, hgbB, ompA, ompH, oma87*, and *sodC*; whereas *tadD, toxA, and pmHAS* genes occurred less frequently. Most of the adhesins and superoxide dismutases and all of the iron acquisition and protectin proteins occurred at significantly (*p*<0.05) higher
frequencies in isolates from pneumonic lungs. Isolates from apparently healthy lungs did not carry the following genes; \textit{hsf-1, hsf-2, tadD, toxA, nanB, nanH}, and \textit{pmHAS}. One adhesion (\textit{hsf-1}) and two iron acquisition (\textit{exbD} and \textit{tonB}) genes occurred at significantly ($p < 0.05$) higher frequencies among \textit{capA} isolates.

Sarangi \textit{et al.} (2014) studied 108 \textit{P. multocida} isolates recovered from various host animals such as cattle, buffalo, swine, poultry (chicken, duck, and emu) and rabbits, which were screened for 8 virulence associated genes. The results revealed some unique information on the prevalence of virulence associated genes among Indian isolates. With the exception of \textit{toxA} gene, all other virulence associated genes were found to be regularly distributed among host species. Association study between capsule type and virulence genes suggested that \textit{pfhA}, \textit{nanB}, and \textit{nanH} genes were regularly distributed among all serotypes with the exception of \textit{capD}, whereas \textit{toxA} gene was found to be positively associated with \textit{capD} and \textit{capA}. The frequency of \textit{hgbA} and \textit{nanH} genes among swine isolates of Indian origin was found to be less in comparison to its equivalents around the globe. Very high prevalence of \textit{tbpA} gene was observed among poultry, swine and rabbit isolates. Likewise, very high prevalence of \textit{pfhA} gene (95.3\%) was observed among Indian isolates, irrespective of host species origin.

Sarangi \textit{et al.} (2015) studied a total of 88 \textit{P. multocida} isolates of small ruminant origin, which were subjected to virulence gene profiling for 19 genes by PCR and antibiogram study employing 17 different antibiotics. Virulence genes like \textit{exbB, exbD, tonB, oma87, sodA, sodC, nanB} and \textit{plpB} (100\% prevalence) and \textit{ptfA} and \textit{hsf-2} (>90\% prevalence) were found to be uniformly distributed among the isolates. A very high prevalence (95.45\%) of \textit{pfhA} gene was observed in the present study. Dermonecrotxin gene (\textit{toxA}) was observed in 48.9\% of the isolates with
highest occurrence among serotype A isolates and interestingly, one of each isolate of serotype B and F were found to carry this gene.

2.2 WHOLE GENOME SHOTGUN SEQUENCING

The basic methodology used for sequencing a large genome is double-barreled shotgun sequencing introduced by Sanger and colleagues (Sanger et al., 1977b). Shotgun sequencing was initially applied by Sanger to the genome of bacteriophage λ (Sanger et al., 1980; Sanger et al., 1982). Over the years, shotgun sequencing has been applied to genomic segments of increasing length. In the 1980s, it was applied to plasmids, chloroplasts (Ohyama et al., 1986), viruses (Goebel et al., 1990), and cosmid clones of length roughly 40 kbp, which were considered to be the limit of this approach. In the 1990s, mitochondria (Oda et al., 1992), bacterial artificial chromosomes of length 200 Kbp carrying genomic DNA from human or other organisms, as well as the entire 1,800 Kbp genome of the bacterium H. influenzae (Fleischmann et al., 1995), were sequenced.

Compared to clone-based shotgun sequencing, Whole Genome Sequencing (WGS) poses a much harder computational challenge, especially in large, repeat-rich genomes. WGS assembly became possible through advances in algorithms and software systems, rather than in either computer processor speed or sequencing accuracy.

The most common technique employed for modern genome projects is whole genome shotgun sequencing (Pittendrigh et al., 2005). Sequencing DNA, sometimes called "reading", involves using chemical reactions to find the exact sequence of nucleotides that make up that molecule. Current laboratory technologies only permit reading sequences of less than 1,000 nucleotides at a time. Genomes of interest are
typically millions or even billions of nucleotides long. In order to determine the genome sequence, it must first be segmented into smaller pieces called fragments. First, many copies of the entire genome are collected and randomly cut. Next, the fragments are sorted by length and groups of fragments with desired sizes are selected. A group of fragments whose approximate sizes are known is called a library. The fragments are placed into a cloning vector such as the bacteria E. coli, which makes many copies of the sequence. Finally, the first 500-1000 nucleotides of each fragment are read using chemical reactions. The result consists of thousands or millions of segments, each 500-1000 nucleotides long, that are known to come from somewhere in the genome.

Figure 2.1: Main steps required for the whole-genome shotgun sequencing of bacterium and generation of genomic data (Fraser and Fleischmann, 1997).
2.2.1 DNA sequencing technologies

The first DNA genome, that of the 5,386-nucleotide, single-stranded bacteriophage \( \phi X174 \) was determined in 1977 using one of the technologies of DNA sequencing invented at the time (Sanger et al., 1977). Since then, sequencing of whole genomes as well as of individual genomic regions and genes has become a major focus of modern biology and transformed the field of modern genetics. However, in the 1970s and for almost another decade, DNA sequencing was barely automated and therefore very tedious process which allowed determining only a few hundred nucleotides in an experiment.

In the late 1980s, semi-automated sequencers with higher throughput became available (Swerdlow and Gesteland, 1990), still only able to determine a few sequences at a time. One breakthrough in the early 1990s was the development of capillary array electrophoresis and appropriate detection systems (Zagursky and McCormick, 1990; Kambara and Takahashi, 1993; Ueno and Yeung, 1994; Kim et al., 1996).

In 1998, the MegaBACE 1000 and the ABI Prism 3700 DNA Analyzer became the first commercial 96 capillary sequencers, a development which then was termed high-throughput sequencing. Only within the last few years, alternative sequencing strategies like pyrosequencing (Ronaghi et al., 1996; Margulies et al., 2005), reversible terminator chemistry (Bentley et al., 2008; Turcatti et al., 2008), sequencing-by-ligation (Shendure et al., 2005), virtual terminator chemistry (Harris et al., 2008) and real-time sequencing (Korlach et al., 2008) were developed or converged into new instruments. These new instruments require us to completely redefine the term "high-throughput sequencing", as they outperform the older Sanger-sequencing technologies by a factor of 100 to 1,000 in daily throughput and reduce
the cost of sequencing one million nucleotides (1Mb) to 4% - 0.1% of that associated with Sanger sequencing. This large difference in throughput led scientists and companies to introduce new terms like "next generation sequencing" (Ansorge, 2009; Metzker, 2009) or "ultra-high-throughput sequencing" (Fox et al., 2009) for this group of new technologies.

2.2.2 Next generation sequencing technologies

Next generation sequencing (NGS) technologies include a number of steps that can be broadly grouped as template preparation, clonal amplification, sequencing, data acquisition and analysis. The unique combination of specific protocols distinguishes one technology from another and determines the type of data produced from each platform. These differences in data output present challenges when comparing platforms based on data quality and cost. Although quality scores and accuracy estimates are provided by each manufacturer, there is no consensus that a 'quality base' from one platform is equivalent to that from another platform (Metzker, 2010). NGS platforms available to dates are listed below and discussed further with particular emphasis on Roche/454's GS-FLX Titanium.

- Roche/454's GS-FLX Titanium
- Illumina/ Solexa's Genome Analyzer
- Life/ APG's SOLiD (Support Oligonucleotide Ligation Detection)
- Polonator G.007
- Helicos Biosciences - Heliscope
- Single Molecule Real Time Sequencing of Pacific Biosciences - SMRT
- Life Technologies' Ion Torrent PGM/ Protone
- Other upcoming sequencing technologies
2.2.2.1 Roche/454's GS-FLX Titanium

The first large-scale adaption of the pyrosequencing technique, invented by 454 Life Sciences (Patrick, 2007) and later commercialized by Roche (Morozova and Marra, 2008), is a high-throughput system. The 454 Genome Sequencer (GS) is based on the pyrosequencing approach developed by Paul Nyren and Mostafa Ronaghi at the Royal Institute of Technology, Stockholm in 1996 (Ronaghi et al., 1996). In contrast to the Sanger technology, pyrosequencing is based on iteratively complementing single strands and simultaneously reading out the signal emitted from the nucleotide being incorporated (also called "sequencing by synthesis" or "sequencing during extension"). Electrophoresis is therefore no longer required to generate an ordered read out of the nucleotides, as the read out is done simultaneously with the sequence extension.

In the pyrosequencing process, one nucleotide at a time is washed over several copies of the sequence to be determined, causing polymerases to incorporate the nucleotide if it is complementary to the template strand. The incorporation stops if the longest possible stretch of complementary nucleotides has been synthesized by the polymerase. In the process of incorporation, one pyrophosphate per nucleotide is released and converted to adenosine triphosphate (ATP) by an ATP sulfurylase. The ATP drives the light reaction of luciferases present and the emitted light signal is measured. To prevent the deoxyadenosine triphosphate (dATP) provided in a typical sequencing reaction from being used directly in the light reaction, deoxy-adenosine-5'-alpha-thio-triphosphate (dATPαS), which is not a substrate of the luciferase, is used for the base incorporation reaction of adenine. Standard deoxyribose nucleotides are used for all other nucleotides. After capturing the light intensity, the remaining
unincorporated nucleotides are washed away and the next nucleotide is provided (Kircher et al., 2009).

Figure 2.2: The pyrosequencing process. One of four nucleotides is washed sequentially over copies of the sequence to be determined, causing polymerases to incorporate complementary nucleotides. The incorporation stops if the longest possible stretch of the available nucleotide has been synthesized. In the process of incorporation, one pyrophosphate per nucleotide is released and converted to ATP by an ATP sulfurylase. The ATP drives the light reaction of luciferases present and a light signal proportional (within limits) to the number of nucleotide incorporations can be measured.

Later pyrosequencing technology was parallelized on a picotiter plate allowing high-throughput sequencing (Margulies et al., 2005). The sequencing plate has about two million wells - each of them able to accommodate exactly one 28μm diameter bead covered with single stranded copies of the sequence to be determined. The beads
are incubated with a polymerase and single-strand binding proteins and together with smaller beads carrying the ATP sulfurylases and luciferases, gravitationally deposited in the wells. Free nucleotides are then washed over the sequencing plate and the light emitted during the incorporation is captured for all wells in parallel using a high resolution CCD (Charged-couple device) camera, exploiting the light-transporting features of the plate used.

One of the main prerequisites for applying this array-based pyrosequencing approach is covering individual beads with multiple copies of the same molecule. This is done by first creating sequencing libraries in which every individual molecule gets two different adapter sequences, one at the 5' end and one at the 3' end of the molecule. In the case of the 454/Roche sequencing library preparation (Margulies et al., 2005), this is done by sequential ligation of two pre-synthesized oligos. One of the adapters added is complementary to oligonucleotides on the sequencing beads and thus allows molecules to be bound to the beads by hybridization.

Low molecule to bead ratios and amplification from the hybridized double stranded sequence on the beads (kept separate using polymerase chain reaction in an water-in-oil emulsion, i.e emulsion PCR) makes it possible to grow beads with thousands of bound copies of a single starting molecule. Using the second adapter, beads covered with molecules can be separated from empty beads (using capture beads with oligonucleotides complementary to the second adapter) and are then used in the sequencing reaction.

The average substitution (excluding insertion/deletions) error rate is in the range of $10^{-3}$ to $10^{-4}$ (Margulies et al., 2005; Quinlan et al., 2008), which is higher than the rates observed for Sanger sequencing but is the lowest average substitution
error rate of the new sequencing technologies. However in Sanger sequencing, *in vitro* amplifications performed for the sequencing preparation causes a higher background error rate, that is, the error introduced into the sample before it enters the sequencing process. In addition, in bead preparation (i.e. emulsion PCR step), a fraction of the beads end up carrying copies of multiple different sequences. These "mixed beads" will participate in a high number of incorporations per flow cycle, resulting in sequencing reads that do not reflect real molecules. Most of these reads are automatically filtered during the software post-processing of the data. The filtering of mixed beads may however cause a depletion of real sequences with a high fraction of incorporations per flow cycle.

A large fraction of the errors observed for GS FLX are small insertions or deletions (InDels), mostly arising from inaccurate calling of homopolymer length, and single basepair deletions or insertions caused by signal-to-noise thresholding issues (Quinlan *et al.*, 2008). Most of these problems can be resolved by higher coverage (Wicker *et al.*, 2006; Green *et al.*, 2008; Quinlan *et al.*, 2008).

As for Sanger sequencing, the error rate increases with the position in the sequence. In the case of 454 sequencing, this is caused by 1) a reduction in enzyme efficiency or loss of enzymes which results in a reduction of the signal intensities, 2) some molecules on the beads no longer being elongated and 3) by an increasing, so-called, phasing effect. Phasing is observed when a population of DNA molecules amplified from the same starting molecule (ensemble) is sequenced, and describes the process whereby not all molecules in the ensemble are extended in every cycle (Kircher and Kelso, 2010).
Current 454/Roche GS FLX Titanium makes it possible to sequence about 1.5 million such beads in a single experiment and to determine sequences of length between 300-500nt. The length of the reads is determined by the number of flow cycles, i.e. the number of times all four nucleotides have been washed over the plate, as well as by the base composition and the order of the bases in the sequence to be determined. Currently, 454/Roche limits this number to 200 flow cycles, resulting in an expected average read length of about 400nt. This is largely due to limitations imposed by the efficiency of polymerases and luciferases which drops over the sequencing run resulting in decreased base qualities.

2.2.2.2 Illumina/ Solexa’s Genome Analyzer

The reversible terminator technology used by the Illumina Genome Analyzer employs the sequencing by synthesis concept that is most similar to that used in Sanger sequencing i.e. the incorporation reaction is stopped after each base, the label of the base incorporated is read out with fluorescent dyes and the sequencing reaction is then continued with the incorporation of the next base (Bentley et al., 2008; Turcatti et al., 2008).

2.2.2.3 Life/ APG’s SOLiD (Support Oligonucleotide Ligation Detection)

The principle behind SOLiD sequencing technology is sequencing-by-ligation, which is very different from the other approaches of pyrosequencing. The sequence extension reaction is not carried out by polymerases but rather by ligases (Shendure et al., 2005). In the sequencing-by-ligation process, a sequencing primer is hybridized to single-stranded copies of the library molecules to be sequenced. A mixture of 8mer probes carrying four distinct fluorescent labels compete for ligation to the sequencing primer. The fluorophore encoding, which is based on the two 3’ most nucleotides of
the probe, is read. Three bases including the dye are cleaved from the 5' end of the
probe, leaving a free 5' phosphate on the extended (by five nucleotides) primer, which
is then available for further ligation. After multiple ligations (typically up to 10
cycles), the synthesized strands are melted and the ligation product is washed away
before a new sequencing primer (shifted by one-nucleotide) is annealed. Starting from
the new sequencing primer, the ligation reaction is repeated.

2.2.2.4 Polonator G.007

The Polonator G.007 is ligation detection sequencing, which decodes the base
by the single-base probe in nonanucleotides (nonamers), not by dual-base coding
(Timp et al., 2010). The fluorophore-tagged nonamers will be degenerated by
selectively ligating onto a series of anchor primers, whose four components are
labeled with one of four fluorophores with the help of T4 DNA ligase, which
correspond to the base type at the query position. In the ligation progress, T4 DNA
ligase is particularly sensitive to mismatches on 3'-side of the gap which is a benefit
to improve the accuracy of sequencing. After imaging, the Polonator chemically strips
the array of annealed primer-fluorescent probe complex; the anchor primer is replaced
and the new mixture of fluorescently tagged nonamers is introduced to sequence the
adjacent base (Deamer and Akeson, 2000).

2.2.2.5 Helicone Biosciences - Heliscope

Helicos is able to sequence individual molecules instead of molecule
ensembles created by an amplification process. Single-molecule sequencing has the
advantage that it is not affected by biases or errors introduced in a library preparation
or amplification step, and may facilitate sequencing of minimal amounts of input
DNA. Using methods able to detect non-standard nucleotides, it could also allow for
the identification of DNA modifications, commonly in the in vitro amplification process (Kircher and Kelso, 2010).

2.2.2.6 Single Molecule Real Time Sequencing of Pacific Biosciences - SMRT

Another technology for sequencing individual molecules is Pacific Biosciences's SMRT (Single Molecule Real Time) sequencing technology (Korlach et al., 2008). This technology performs the sequencing reaction on silicon dioxide chips with a 100nm metal film containing thousands of tens of nanometers diameter holes, so called zero-mode waveguides (ZMWs) (Eid et al., 2009). Each ZMW is used as a nano visualization chamber, providing a detection volume of about 20 zeptoliters ($10^{-21}$ liters). At this volume, a single molecule can be illuminated while excluding other labelled nucleotides in the background - saving time and sequencing chemistry by omitting wash steps. In recent new sequencing technologies like Helicos and Pacific Biosciences systems, the sequencing of single molecules without prior library preparation or amplification is achieved.

2.2.2.7 Life Technologies's Ion Torrent PGM/Protone

A completely different approach to next-generation sequencing is embodied in an instrument system that detects the release of hydrogen ions, a by-product of nucleotide incorporation, as quantitated changes in pH through a novel coupled silicon detector. This instrument was commercialized in 2010 by Ion Torrent (Rothberg et al., 2011). Enriched beads are primed for sequencing by annealing a sequencing primer and are deposited into the wells of an ion chip, a specialized silicon chip designed to detect pH changes within individual wells of the sequencer as the reaction progresses stepwise. Ion chip has an upper surface that serves as a
microfluidic conduit to deliver the reactants needed for the sequencing reaction. The lower surface of the ion chip interfaces directly with a hydrogen ion detector that translates released hydrogen ions from each well into a quantitative readout of nucleotide bases that were incorporated in each reaction step.

2.2.2.8 Other upcoming sequencing technologies

Three other systems viz. Oxford Nanopore's BASE technology (Clarke et al., 2009), IBM's silicon based nanopores (IBM Research, 2009) and Life Technologies single molecule sequencing technology based on quantum dots (Thompson and Milos, 2011) have also been reported.

Oxford Nanopore's BASE technology offers the potential to identify individual nucleotide modifications (e.g. 5-methyl-cytosine versus cytosine) during the sequencing process (Clarke et al., 2009). In Nanopore sequencing, a multilayer metal/dielectric nanopore device is developed which utilizes the interaction of the DNA backbone charges with a modulated electric field to trap and slowly release an individual DNA molecule. For Life Technologies quantum dots, light-emitting semiconductor nanocrystals of 2-10 nm diameter are attached to DNA polymerases. These dots can be laser excited with a specific wavelength and then via fluorescence resonance energy transfer (FRET) enable light emission from the fluorescently labelled nucleotides at a different wavelength while the polymerase incorporates them during complementary strand synthesis. These quantum dots do not only provide/transfer the energy for the fluorescence signal, they also enhance the signal strength.
2.2.3 Whole genome sequencing of *P. multocida*

The first complete genome sequence of a common avian clone of *Pasteurella multocida*, Pm70, was reported by May *et al.* (2001). The genome of Pm70 is a single circular chromosome 2,257,487 base pairs in length and contains 2,014 predicted coding regions, 6 ribosomal RNA operons, and 57 tRNAs. Genome-scale evolutionary analyses based on pairwise comparisons of 1,197 orthologous sequences between *P. multocida*, *Haemophilus influenzae* and *Escherichia coli* suggested that *P. multocida* and *H. influenzae* diverged ≈270 million years ago and the 'Y' subdivision of the proteobacteria radiated about 680 million years ago. Two undescribed open reading frames, accounting for ≈1% of the genome, encode large proteins with homology to the virulence-associated filamentous hemagglutinin of *Bordetella pertussis*. Consistent with the critical role of iron in the survival of many microbial pathogens *in silico* and whole genome microarray analyses identified more than 50 Pm70 genes with a potential role in iron acquisition and metabolism.

Liu *et al.* (2012) completed whole genome sequencing of *Pasteurella multocida* HN06, a Toxigenic Strain of Serogroup D by using the Illumina Solexa platform. A total of 4,792,617 paired-end reads were generated from a genomic DNA library with 300- to 500-bp fragments, giving an ~360-fold coverage of the genome. All reads were *de novo* assembled using the Velvet package, and 61 contigs of more than 500 bp were generated. Gaps between contigs were closed by primer walking. Protein-coding sequences (CDSs) were predicted by use of the programs Glimmer and GeneMarks. tRNAs and rRNAs were identified using tRNAscan-SE and RNAmmer, respectively. *P. multocida* HN06 consists of a circular chromosome with 2,402,218 bp and a plasmid, pHN06, with 5,360 bp. The average GC contents of the chromosome and plasmid are 40.2% and 47.5%, respectively. The HN06 genome
encodes 2,279 CDSs with an average size of 928 bp, accounting for about 88% of the genome. Additionally, 56 tRNA genes and 6 rRNA operons were identified. Approximately 48.4% of CDSs were assigned to functional COGs (cluster of orthologous groups), 3.9% had general function predictions only, and the remaining proteins had unknown functions.

Abrahante et al. (2013) sequenced two virulent strains of Pasteurella multocida, P1059 and X73, using 454 pyrosequencing. The assembly generated 2.30 Mbp (40.21% GC, 24 contigs, 18 contigs >500 bp) and 2.26 Mbp (40.30% GC, 39 contigs, 28 contigs >500 bp) genomes for P1059 and X73, respectively. Genome annotation using the RAST server revealed that P1059 contained 2,144 predicted coding regions, a gene density of 88.8%, an average coding size of 961 bp, 50 tRNAs, and 4 rRNA operons. X73 contained 2,085 predicted coding regions, a gene density of 88.30%, an average coding size of 964 bp, 51 tRNAs, and 4 rRNA operons. Multigenome comparative analysis identified the presence of a filamentous hemagglutinin pfhB1 gene, which was highly conserved in all three strains including Pm70. The pfhB2 gene was identical in X73 and P1059 strains but shared only 90% sequence identity with the Pm70 strain. A novel gene, pfhB3, was present in both X73 and P1059, but absent from Pm70. Another novel gene, pfhB4, was unique to the P1059 strain. The plpE gene was present in all three strains and highly conserved. There was presence of a chondroitin synthase gene (fcbD) in Pm70 and the hyaluronan synthase gene (hyaD) in both P1059 and X73. The pcgDABC gene cluster involved in the decoration of phosphocholine to the outer core of the lipopolysaccharide (LPS) was present only in X73, the pm1138 gene encoding a glycosyltransferase was present in all the three strains.
Eidam et al. (2013) performed whole genome sequencing of *Mannheimia haemolytica* Strain 42548 from a case of bovine respiratory disease using the 454 GS-FLX titanium XL system (Titanium GS70 chemistry) and the genome analyzer (GA) IIx (Illumina). 454 shotgun sequencing produced 135,530 single-end reads with an average length of 612 bases (28.6X coverage), while Illumina sequencing resulted in 3,696,221 of 112-bp paired-end reads (116X coverage). The initial hybrid de novo assembly performed using the MIRA software resulted in 316 contigs. Gaps between contigs were closed using PCR amplification, subsequent to Sanger sequencing. Gap closure was done using the Gap4 (v.4.11) software. The genome of *M. haemolytica* 42548 consists of a single chromosome of 2.73 Mb with a GC content of 41.05%. The initial gene prediction was done using YACOP and Glimmer, while rRNA and tRNA were identified with RNAmmer and tRNAscan. This revealed a total number of 2,888 genes, including 6 rRNA gene clusters and 61 tRNA genes. Protein genes annotated using Swiss-Prot and TrEMBL resulted in 2,807 protein-encoding genes with assigned functions.

Hauglund et al. (2013) carried whole genome sequencing of *Mannheimia haemolytica* strain D153 isolated from pneumonic calf lung using three platforms: the Roche (454) GS FLX titanium, resulting in 26-fold coverage; the Illumina GA IIx, resulting in 1,700-fold coverage; and the pacific bioscience Research (PacBio RS), resulting in 30 fold coverage. The completed D153 genome sequence consists of 2.68 Mb, with a G+C content of 41.04%. The draft genome sequence of *M. haemolytica* D193 was determined using the Roche platform alone, which yielded 39-fold coverage. Assembly against the closed D153 reference genome using the CLC software yielded 50 contigs consisting of 2.68 Mb, with a G+C content of 41.08%, an N50 of 139,865 bp, and 100% of contigs with lengths of >500 bp. Annotation of both
genome sequences was accomplished with the NCBI Prokaryotic Genome Annotation Pipeline. Strain D153 contained a total of 2,766 genes, including 2,641 predicted protein-encoding genes, 43 frame shifted pseudogenes, 20 rRNA genes, and 62 tRNA genes. Strain D193 contained a total of 2,801 genes, including 2,689 predicted protein encoding genes, 39 frameshifted pseudogenes, 16 rRNA genes, and 57 tRNA genes. One CRISPR array was detected in each isolate. In contrast to the multi-resistant *M. haemolytica* isolate 42548, both strains D153 and D193 lack the genes *aphA1*, *strA*, *strB*, and *sul2*. *tetR* and *tetH* were detected in strain D193 but not D153.

Lainson et al. (2013) performed whole genome sequencing for three bovine isolates (2000, R11F, 1500E) from the United Kingdom and one bovine isolate from United States (P1933) using an Illumina Solexa Genome Analyzer. *De novo* assembly was carried out using Velvet version 0.7, and the resulting sequence data were submitted to NCBI analysis using the Prokaryotic Genomes Annotation Pipeline (PGAP). The cumulative sequence lengths and the numbers of predicted coding sequences are consistent with those of related *P. multocida* strains sequenced previously (accession numbers NC_017764, NC_016808, NC_17027, and NC_002663). Assembled contig sequences were mapped to the reference genome Pm70 using Nucmer. This showed extensive sequence similarity, with a high proportion of each genome mapping to Pm70 with an identity of 98% or greater.

Lainson et al. (2013b) isolated *P. multocida* strain 671/90 which is an A:3 serotype isolated from infected calf lung tissue and sequenced using Illumina PE and 454 platforms and produced 26 contigs ranging in size from 124 to 578,571 bp, with a cumulative size of 2,246,063 bp. The G+C content was found to be 40.25% and predicted 2,115 genes, of which 2,064 were protein coding, 5 encoded rRNAs, and 46
encoded tRNA sequences. Strain 671/90 was having extensive homology 95.98%, and mapped with an identity of ≥98% to Pm70 strain.

Yap et al. (2013) sequenced strain PMTB isolated from buffalo carcass, performed using an Illumina genome analyzer. A total of 7,760,284 single-end reads were generated from a genomic library with 300- to 500-bp fragments with approximately 100-fold genome coverage. PMTB has a genome size of 2,203,419 bp, a coding percentage of 87.3%, a G+C content of 40.44%, and 9 unclosed gaps. From the analysis of PGAAP results, strain PMTB contains 2,021 coding sequences (CDS) with an average size of 951bp. Among the 2,021 CDS, 1,796 were assigned to functional clusters of orthologous groups (COGs), 51 had general function prediction only, and the remaining proteins had unknown functions. The toxA gene, which is frequently detected in genomes of serotype D, was confirmed by PCR assay to be absent in the genome of strain PMTB. The integrative conjugative elements of P. multocida (ICEPmul) were not found in the draft genome sequence of strain PMTB. A comparative analysis of the protein sequences encoded by the genome of P. multocida serotype F:3 strain Pm70 and that of PMTB revealed that at 50% identity, 217 protein sequences encoded by strain PMTB represented unique proteins. Of these unique proteins, 22 were phage or transposon related.

Abrahante et al. (2014) performed whole-genome sequencing of two strains, 2213 and 3213 of P. multocida isolated from the blood of carcasses of two buffaloes (Bubalus bubalis) using the Illumina MiSeq platform. A total of 6,253,700 and 5,550,478 paired-end 150-bp reads of each genome were assembled into 34 and 32 contigs, respectively, for strains 2213 and 3213, using Velvet assembly software available on the Galaxy Suite (Minnesota Supercomputing Institute). Major gaps (>50 bp) were resolved on the GenBank assembly algorithm in an iterative manner in order
to obtain a single assembled whole genome sequence. An annotation file was generated by Rapid Annotations using Subsystems Technology (http://rast.nmpdr.org/rast.cgi), and cross-verified using a re-annotation performed by the Genbank annotation resource. RAST annotation of strain 2213 using strain 36950 as a reference produced a full genome of 2,309,333 bp carrying 2,163 coding sequences classified into 425 metabolic or virulence subsystems and 59 predicted RNAs. Strain 3213 produced a genome or 2,307,438 bp carrying 2,160 coding sequences that were classified into 424 subsystems and 57 predicted RNAs.

Davenport et al. (2014) performed sequencing of P. multocida subsp. multocida ATCC 43137, a type-A LPS strain commonly used as a reference strain in pathogenicity studies using Illumina technology. Genome assembly was performed by the Los Alamos National Laboratory (LANL) Genome Science Group, and 300-fold 100-bp paired-end (270 ± 30 bp insert) Illumina data assembled in Newbler (version 2.6), Velvet (version 1.2.08), and AllPaths (version 42298). Consensus sequences from all assemblers were computationally shredded and assembled with a subset of read pairs from the long-insert library using Phrap (version SPS-4.24). The resulting assembly was brought to closed and finished status through both manual and computational finishing efforts using Consed and in-house scripts. The assembled genome sequence was corrected by mapping Illumina reads back to the final consensus sequences using Burrows-Wheeler Alignment (BWA), SAMtools and inhouse scripts. Annotations were completed at LANL using an automated system utilizing the ergatis workflow manager and in-house scripts. The 2.27 Mbp (40.4% G + C content) complete assembly of P. multocida subsp. multocida ATCC 43137 assembly includes 2,076 coding sequences, 19 rRNA, and 58 tRNA sequences.
Preliminary review of the annotated genome indicates resistance genes for multiple toxic metals, iron acquisition, hemagglutinin, and genes for LPS production.

Vaid et al. (2014) sequenced *P. multocida* isolate strain VTCCBAA264, a B:2 serotype, from buffalo intestine using 454 pyrosequencing. A total of 123,415 reads were with ~23X coverage of the genome with 78 contigs ranging in size from 710 to 319,560 bp, with a total size of 2,280,332 bp. Annotation using RAST server showed G+C content of 40.4% with 2,176 predicted genes, out of which, there were 2,127 protein-encoding genes, 4 coded rRNAs, and 45 encoded tRNAs. Among the 2,127 coding sequences (CDSs), 2,074 (97.5%) were assigned to functional clusters of orthologous groups (COGs), 1,863 (87.6%) were assigned to FIGfams (fellowship for interpretation of genome families) and 1,527 (71.8%) to KEGG (Kyoto Encyclopedia of Genes and Genomes). One clustered regularly interspaced short palindromic repeats (CRISPRs) array was also detected. The genes for antimicrobial resistance including translation elongation factor, DNA gyrase subunit A, topoisomerase IV subunit B, and DNA directed RNA polymerase subunit, were found among others. Genes for resistance to fluoroquinolones and negative regulators of beta-lactamase expression were also reported.

### 2.3 SEQUENCE ANALYSIS STRATEGIES

Genome sequencing presents challenges to computer analysis at all levels, from sequence assembly to large-scale genome comparisons. Genome sequence analysis is a multistep process that starts with establishing the maximally accurate assembled sequence, and proceeds through functional predictions to higher-level genome comparisons (Koonin et al., 1996).
Figure 2.3: Whole genome shotgun sequencing. The obtained contigs are ordered into scaffolds using information from read pairs spanning the gaps.

2.3.1 Bacterial genome assembly

The process of deciphering the sequence of a genome from the small DNA fragments and any other additional genome information available is called "assembling" the genome. At present, the sequencing process is often a 10x-fold genome coverage followed by two phases: assembly and finishing. Currently, sequencing reads generated by different sequencing technologies range from 35-1000 bp. In order to obtain a complete bacterial genome, the fragments need to be aligned and joined together using a computer program called an assembler. Assemblers can join sequences together based on overlapping regions between the sequences, assuming that the two sequence reads have originated from the same place in the genome. After assembly, a collection of contiguous pieces (contigs) instead of an entire chromosome are usually obtained. This is often due to non-random shearing of DNA, intrinsic cloning bias and repeated regions in the genome. Increasing the sequencing coverage of the genome will help to reduce the number of contigs. Lander and Waterman (1988) showed that sequencing a 1 Mbp genome using Sanger
chemistry resulted in a small number of contigs (~5) if 8 to 10 times genome coverage was attained.

In addition to the assembler, another computer program called a scaffolder is used if paired-end or mate-pair reads are available. Scaffolder can link distant sequences together based on the distance between the two ends of the original template, i.e. 3 kb or 8 kb. Therefore, the scaffolder is able to define the size of gaps between contigs and orient the contigs into a draft genome (Pop et al., 2004). To assemble next-generation sequencing reads, de novo assembly becomes more of a challenge because only short overlaps can be considered. Therefore, higher coverage of the genome is required for assembly of shorter reads, resulting in large volumes of data (Miller et al., 2010). However, one study has shown that it is possible to assemble a large portion of the E. coli K12 MG1655 genome using read lengths of 20-50 nucleotides, given that the reads were error-free (Whiteford et al., 2005). Since then, several bacterial genomes have been de novo assembled using short read sequence data, including Helicobacter acinonychis (Dohm et al., 2007; Hernandez et al., 2008), Staphylococcus aureus (Hernandez et al., 2008), Bacillus subtilis (Srivatsan et al., 2008; Nishito et al., 2010), Pseudomonas aeruginosa (Salzberg et al., 2008), Pseudomonas syringae (Reinhardt et al., 2009; Studholme et al., 2009) and Erwinia pyrifoliae (Smits et al., 2010). Furthermore, combining the read data from more than one type of sequencing platform can improve de novo assembly of a bacterial genome significantly (Aury et al., 2008; Reinhardt et al., 2009, Smits et al., 2010).

2.3.1.1 Construction of contigs and scaffolds

The placement of the reads along a reference genome implicitly defines a set of contigs or contiguous regions of the assembly, as well as the relative order and
orientation of these contigs in a structure which is commonly known as a scaffold (Pop et al., 2004). This phase consists of three major steps. Firstly, each overlap is evaluated based on the depth of coverage of two regions in the overlap. Secondly, poorly differentiated ends of every sequence read are identified and trimmed. Thirdly, reads are assembled into contigs based on unique overlaps.

Contigs are corrected and linked into scaffolds based on pairs, and scaffolds are also corrected based on read pairs. The joining of the fragments is modeled as a mathematical weighted graph, where nodes are fragments and the weights of edges are the number of overlapping nucleotides. The fragments are joined based upon maximum overlap using a greedy algorithm (Qin et al., 2003). In a greedy algorithm, most nodes having maximum (or minimum) scores are collapsed first.

To join contigs, the fragments with larger nucleotide sequence overlaps are joined first. Contigs are constructed by processing overlaps with an adjusted score greater than a cutoff. Initially, each read is a contig by itself. The overlaps are ranked in a decreasing order of their adjusted scores, and they are considered one by one, in order, for the construction of contigs. The overlap being considered is called the current overlap. For a current overlap between two reads, if the reads are in different contigs and two contigs have an overlap consistent with the current overlap, the two contigs are merged into a larger contig.

The computation is performed on one processor with enough memory to hold the overlaps and contigs. Read pairs are used to order and orient contigs into a scaffold, as follows. Initially, each contig is a scaffold by itself. Unsatisfied read pairs are partitioned into groups such that all reads in a group link a pair of scaffolds. The groups of unsatisfied read pairs are considered in a decreasing order of their sizes. If
the number of read pairs in the group is sufficiently large, the read pairs link two scaffolds, and the two scaffolds can be combined by using the read pairs in the group, then the two scaffolds are combined into an even larger scaffold.

2.3.1.2 Generation of consensus sequences

For each group of overlapping reads in the refined layout, a multiple alignment is computed to generate a consensus sequence for the genomic region covered by those reads. The multiple alignment is computed in a series of rounds. In each round, a pairwise alignment of each read to the current consensus sequence is computed and the resulting multiple alignment is used to generate a new consensus sequence. The process terminates when the new consensus sequence is the same as the one in the previous round. This essentially utilises the algorithm described by Anson and Myers (1997). For each scaffold, a set of repetitive reads that are linked by read pairs to unique reads in the scaffold are identified. For each gap in the scaffold, a subset of repetitive reads that may fall into the gap are selected from the set on the scaffold. An attempt is made to close the gap with the subset of repetitive reads. After all gaps in the group of scaffolds are considered for closure, a consensus sequence is generated for each contig and a list order and oriented contig consensus sequence is recorded for each scaffold.

Generation of a consensus sequence for each contig is based on multiple alignments of reads in the contig, which is constructed as follows. The reads in the contig are sorted in an increasing order of their position in the contig. A multiple alignment is constructed by repeatedly aligning the current read with the current alignment, and the resulting alignment is the current alignment for the iteration. The reads in the contig are considered one by one, in order. For each column of the final
multiple alignment, a weighted sum of quality values is calculated for base type. The base type with the largest sum of quality value is taken as the consensus base for the column.

2.3.1.3 Scaffold and gap closure

The availability of complete edited genome sequences of bacterial pathogens free of gaps is of great utility as it gives access to complete gene sets and provides for studies of genome organisation, genome comparisons and functional genomics, including the development of microarrays (Fraser et al., 2000). Once one or two genomes of a given organism are sequenced to completion, the diversity of the species can be assessed through comparative genome hybridisation (CGH) using microarrays (Tettelin et al., 2001), or by generating draft sequences of other strains of interest and comparing them to the reference genome(s). These two approaches are rapid and cost effective whilst generating genome-scale information on species diversity.

2.3.2 Bacterial genome annotation

In general, genome annotation can be performed at two main levels, static and dynamic (Medigue and Moszer, 2007). At the static level, the main features of the genome can be obtained, such as protein coding genes, functional RNA products, GC content, codon usage, genomic islands, motifs, chemical and structural properties of proteins and their sub-cellular localisation. In contrast, the dynamic view can exhibit gene context, gene order, regulatory networks, protein interaction networks, metabolic networks as well as phyloprofile and gene fusion/fission information obtained by comparative genomics (Medigue and Moszer, 2007). Nowadays, bacterial genomes are often annotated using automated pipelines, which can be either web-based or run locally (Stothard and Wishart, 2006). Web-based annotation pipelines are more
convenient for small research groups or small sequencing facilities that lack computing resources and expertise that is necessary to maintain or implement the software (Stothard and Wishart, 2006). Bacterial Annotation System (BASys) is an automated bacterial annotation web server (http://wishart.biology.ualberta.ca/basys), which uses more than 30 programs to determine nearly 60 annotation subfields for each gene (Van Domselaar et al., 2005). With this system, results can be generated in about 24 h for a 5 megabase genome and can be browsed and evaluated using a navigable graphical map. However, BASys is not able to analyze partially assembled genomes (Van Domselaar et al., 2005).

In 2008, the RAST (Rapid Annotation using Subsystem Technology) Server (http://rast.nmpdr.org/), a fully automated annotation service for complete or draft archaeal and bacterial genome was built (Aziz et al., 2008). Annotations provided by the service include protein-encoding, rRNA and tRNA genes, gene function and metabolic network. On the completion of annotation, the annotated genome can be downloaded in a variety of formats or browsed in SEED Viewer for up to 120 days (Aziz et al., 2008). The SEED is an annotation/analysis tool provided by the Fellowship for Interpretation of Genomes (FIG) (Overbeek et al., 2005). WeGAS (http://ns.smallsoft.co.kr:8051) is another web-based microbial genome annotation system (Lee et al., 2009). Like the RAST Server, it is capable of handling both ongoing and completed microbial genome projects. The user can start genome annotation with contigs and the process can be monitored during each analysis. The annotation pipeline includes seven major modules, which are gene prediction, homology search, promoter search, pathway mapping, motif search, COG (Clusters of Orthologous Groups of proteins) assignment and GO (Gene Ontology) assignment. A genome browser is also available to view the detailed results (Lee et al., 2009).
Recently, another prokaryotic genome annotation web server called Integrative Services for Genomics Analysis (ISGA) became available to the community (http://isga.cgb.indiana.edu/) (Hemmerich et al., 2010).

2.3.2.1 Genome annotation pipeline

Genome annotation falls into two distinct stages (Kuroda and Hiramatsu, 2004). The first stage is referred to as 'structural annotation' and involves the correct identification and localisation of distinct sequence elements such as genes, regulatory elements, transposons, repetitive elements and more. The second stage, termed 'functional annotation' attempts to predict the biological function for each of those elements and the biological process in which it takes part.

2.3.2.1.1 Structural annotation

Computational identification of protein-coding genes is an approach used in newly sequenced genomes. It is divided into three methods: (i) ab initio or de novo methods, which predict genes solely on the basis of local sequence characteristics (ii) similarity-based methods, which utilise sequence similarity to known genes and (iii) comparative methods, which employ sequence comparison between multiple, related genomes to identify conserved genes. Gene prediction is the most visible part of this phase and involves identification of the genes encoded in the sequence, which include the boundaries of the regions that act as templates for transcription, the initiator and terminator of translation, splice sites, promoter and regulatory regions and perhaps several other biologically important elements. In bacterial genomes, gene prediction is largely a matter of identifying long ORFs (open reading frame). For a known sequence, ORFs are predicted using a variety of programs such as GeneMark (Besemer et al., 2001), GeneMark.hmm (Lukashin and Borodovsky, 1998),
GENESCAN (McEvoy et al., 1998), Glimmer (Aggarwal and Ramaswamy, 2002) and Glimmer3 (Delcher et al., 2007). Each program applies to the sequence for searching the potential ORFs or gene encoding regions. A genome can be divided into two parts, one comprises the protein and RNA encoding genes and the other is the non-coding DNA (Pearson and Lipman, 1988).

2.3.2.1.2 Functional annotation

Once the genes and other structural sequences in a genome have been identified, the next stage in annotation is to predict the molecular function and biological role of these elements. Functional annotation is the process of assigning function to the genes expressed in the genome, discovering the mechanisms of regulation associated with different regulatory sites and more generally, assigning function to structural elements annotated in the genome. Functional annotation is divided into two parts, functional annotation based on homology or based on protein signature.

Homology-based function prediction involves the use of database searches to identify genes that are similar to a query sequence and which preferably have a known experimentally determined function. Similar to structural annotation, the focus in functional annotation is on the function of genes and their products. A widely used tool for homology searches in databases is the Basic Local Alignment Search Tool (BLAST) algorithm (Altschul et al., 1997). BLAST is based on a search and alignment algorithm and is used to compare a nucleotide or protein sequence against nucleotide or protein databases. Similarity scores between the query sequence and database homologous reflect their local pairwise alignments. Homology searches are carried out on the predicted ORFs in an attempt to determine function as well as other
information about the potential gene and its protein product (Rubin, 2001). Other sequence similarity search tools such as FASTA (Pearson and Lipman, 1988) and variations of BLAST are commonly used algorithms for sequence comparisons (Altschul et al., 1997). FASTA generates tables of short query sequences to compare to the database and BLAST expands upon the short matches to find the best scoring matches for the query sequence. The possible ORFs are classified by comparison to the entire sequence. Many different databases are available for searching including the NCBI non-redundant database (NCBI nr) (Altschul et al., 1997), the KEGG database for examining metabolic pathways (the Kyoto Encyclopedia of Genes and Genomes, Japan) (Kanehisa et al., 2008), the Clusters of Orthologous Groups (COG) (Tatusov et al., 2003) and the STRING-extended COG database (179 microbial genomes, version 7.0) (von Mering et al., 2007).

The limitation of homology-based function prediction, when tools such as BLAST are used, is that the presence of structural domains and motifs that make up a protein are not properly analysed. BLAST may report the most significant database matches for a query sequence that are based solely on the presence of one common, conserved protein domain. Other domains in the sequence may be different and be indicative of an alternative function.

2.3.3 Data analysis

The process of sorting quality reads and aligning and analyzing hundreds of thousands or millions of base pairs is both time intensive and computationally intensive. Technology for data acquisition is proceeding faster than information technology in many cases and data-processing time may well exceed sample-handling time until new methods of analysis are developed.
Figure 2.4: Workflow for high-throughput whole genome sequencing in bacteria

**Explanation of Figure 2.4: Sample collection:** A biological sample (e.g., blood) is collected. **Culture:** Bacterial colonies are isolated from the sample by culturing on appropriate media. **DNA Preparation:** DNA is extracted from the colonies and a DNA library is prepared ready for sequencing. **High-Throughput Sequencing:** Millions of short sequence reads are yielded, typically several hundred nucleotides long or less. To reconstruct the genome, one of two approaches is generally adopted. **Mapping to Reference Genome:** In reference-based mapping, the short sequences are mapped (i.e., aligned) to a reference genome using an algorithm. Preferably, the reference genome is of high quality, complete, and closely related. The pie chart illustrates that not all reads necessarily map to the reference genome (e.g., because of novel regions not present in the reference). **Filtering:** Short reads cannot...
be mapped reliably to repetitive regions of the reference genome, so these are identified and filtered out. Sites that are problematic for other reasons (e.g., because too few reads have mapped or because the consensus nucleotide is ambiguous) are also filtered out. The pie chart illustrates that some portion of the reference genome does not get called due to filtering. In the mapped genome, these positions will receive an ambiguity code (i.e., N rather than A, C, G, or T). **De novo Assembly of Contigs:** An alternative to mapping is *de novo* assembly, in which no reference genome is used. An algorithm is used to assemble short reads into longer sequences known as contigs. The number and length of contigs will depend on general factors such as the length of sequence reads and the total amount of DNA sequence produced, as well as local factors such as the presence of repetitive regions. The pie chart shows an example of the proportion of all reads that assemble into contigs of a given length. **Alignment:** For further analysis, it is necessary to align local regions (e.g., genes) or whole genomes using appropriate algorithms. There is a trade-off in computational terms between the length of region and the number of sequences that can be aligned. **Sequence Analysis:** The two approaches produce sequence alignments that represent pairwise alignments against a reference (mapping) or multiple alignments one to another (*De novo* assembly). These alignments can be analyzed directly, or processed further to detect variants such as single nucleotide polymorphisms, insertions, and deletions (Wilson, 2012). Various bioinformatics tools/software and the useful databases at present, are shown in Table 2.2
<table>
<thead>
<tr>
<th>Tool</th>
<th>Description/ Features</th>
<th>Web address/URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abyss</td>
<td>Alignment/assembly/Illumina</td>
<td><a href="http://www.bcgsc.ca/platform/biolinfo/software/abyss">www.bcgsc.ca/platform/biolinfo/software/abyss</a></td>
</tr>
<tr>
<td>Phrap</td>
<td>Alignment/assembly/Illumina/Roche</td>
<td><a href="http://www.phrap.org/conshed/conshed.html">http://www.phrap.org/conshed/conshed.html</a></td>
</tr>
<tr>
<td>Velvet</td>
<td>Alignment/assembly/Roche/ABI/Illumina</td>
<td><a href="http://www.ebi.ac.uk/%7Ezzerbino/velvet">http://www.ebi.ac.uk/%7Ezzerbino/velvet</a></td>
</tr>
<tr>
<td></td>
<td><strong>Assembly tools</strong></td>
<td></td>
</tr>
<tr>
<td>Glimmer</td>
<td>Microbial gene-finding system</td>
<td><a href="http://www.cbcb.umd.edu/software/glimmer/">www.cbcb.umd.edu/software/glimmer/</a></td>
</tr>
<tr>
<td>GeneMark</td>
<td>Gene prediction in bacteria, archaea and metagenomes</td>
<td><a href="http://opal.biology.gatech.edu/GeneMark/">http://opal.biology.gatech.edu/GeneMark/</a></td>
</tr>
<tr>
<td>EasyGene</td>
<td>Gene predictor in prokaryotic DNA</td>
<td><a href="http://www.cbs.dtu.dk/services/EasyGene/">www.cbs.dtu.dk/services/EasyGene/</a></td>
</tr>
<tr>
<td>FgenesB</td>
<td>Bacterial operon and gene prediction</td>
<td><a href="http://linux1.softberry.com/">http://linux1.softberry.com/</a></td>
</tr>
<tr>
<td>REGANOR</td>
<td>Gene prediction sever and database</td>
<td><a href="http://www.cebitec.uni-bielefeld.de/">www.cebitec.uni-bielefeld.de/</a></td>
</tr>
<tr>
<td>Prodigal</td>
<td>Prokaryotic dynamic programming gene finding algorithm</td>
<td><a href="http://prodigalornl.gov/">http://prodigalornl.gov/</a></td>
</tr>
<tr>
<td></td>
<td><strong>Gene prediction tools</strong></td>
<td></td>
</tr>
<tr>
<td>GenColors</td>
<td>Comparative genomics and annotation platform</td>
<td><a href="http://www.genoscope.cns.fr/">http://www.genoscope.cns.fr/</a></td>
</tr>
<tr>
<td>MicroScope</td>
<td>Comparative genomics and annotation platform me</td>
<td><a href="http://www.genoscope.cns.fr/">http://www.genoscope.cns.fr/</a></td>
</tr>
<tr>
<td>KAAS</td>
<td>KEGG automatic annotation server</td>
<td><a href="http://www.genome.jp/tools/kaas/">www.genome.jp/tools/kaas/</a></td>
</tr>
<tr>
<td>AutoFACT</td>
<td>Automated annotation tool</td>
<td><a href="http://megasun.bch.umontreal.ca/">http://megasun.bch.umontreal.ca/</a></td>
</tr>
<tr>
<td>BASys</td>
<td>Bacterial annotation system</td>
<td><a href="http://basys.ca/basys/cgi">http://basys.ca/basys/cgi</a> submits.pl</td>
</tr>
<tr>
<td>IGS</td>
<td>IGS prokaryotic annotation pipeline</td>
<td><a href="http://ae.igs.umaryland.edu/cgi/">http://ae.igs.umaryland.edu/cgi/</a></td>
</tr>
<tr>
<td>CMR</td>
<td>Comprehensive microbial resource and annotation</td>
<td><a href="http://cmr.jcvi.org/">http://cmr.jcvi.org/</a></td>
</tr>
<tr>
<td>GenDB</td>
<td>Prokaryotic genomes annotation system</td>
<td><a href="http://www.cebitec.uni-bielefeld.de/">www.cebitec.uni-bielefeld.de/</a></td>
</tr>
<tr>
<td>HAMAP</td>
<td>Automated and manual annotation of microbial proteomes</td>
<td><a href="http://us.expasy.org/sprot/hmap/">http://us.expasy.org/sprot/hmap/</a></td>
</tr>
<tr>
<td>RAST</td>
<td>Rapid annotation using subsystem technology</td>
<td><a href="http://www.nmpdr.org/FIG/wiki/view/cgi/">www.nmpdr.org/FIG/wiki/view/cgi/</a></td>
</tr>
<tr>
<td>xBASE</td>
<td>Bacterial genome annotation service</td>
<td><a href="http://www.xbase.ac.uk/annotation/">http://www.xbase.ac.uk/annotation/</a></td>
</tr>
<tr>
<td>Blast2GO</td>
<td>Annotation and sequence analysis tool</td>
<td><a href="http://www.blast2go.com/">http://www.blast2go.com/</a></td>
</tr>
<tr>
<td></td>
<td><strong>Automatic and manual annotation pipelines/Tools</strong></td>
<td></td>
</tr>
<tr>
<td>Databases and resources</td>
<td>Description</td>
<td>URL</td>
</tr>
<tr>
<td>--------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>EMBL</td>
<td>Nucleotide sequence database</td>
<td><a href="http://www.ebi.ac.uk/embl">http://www.ebi.ac.uk/embl</a></td>
</tr>
<tr>
<td>GOLD</td>
<td>Data resource for genomic and matagenomic projects</td>
<td><a href="http://www.genomesonline.org/">http://www.genomesonline.org/</a></td>
</tr>
<tr>
<td>KEGG</td>
<td>An integrated database resource, provides genomic, chemical and systemic information</td>
<td><a href="http://www.kegg.jp/">http://www.kegg.jp/</a></td>
</tr>
<tr>
<td>IMG</td>
<td>Resource for comparative analysis and annotation</td>
<td><a href="http://img.jgi.doe.gov/">http://img.jgi.doe.gov/</a></td>
</tr>
<tr>
<td>JCVI</td>
<td>Comprehensive microbial resource (CMR)</td>
<td><a href="http://www.jcvi.org/">http://www.jcvi.org/</a></td>
</tr>
<tr>
<td>MBGD</td>
<td>Database analysis of orthologous, paralogous, motifs, gene order and annotation</td>
<td><a href="http://mbgd.genome.ad.jp/">http://mbgd.genome.ad.jp/</a></td>
</tr>
<tr>
<td>RDP</td>
<td>Ribosomal database, bacterial RNA sequences, alignments and tools for RNA analysis</td>
<td><a href="http://rdp.cme.msu.edu/">http://rdp.cme.msu.edu/</a></td>
</tr>
<tr>
<td>Rfam</td>
<td>RNA database</td>
<td><a href="http://rfam.sanger.ac.uk/">http://rfam.sanger.ac.uk/</a></td>
</tr>
<tr>
<td>GtRNAadb</td>
<td>RNA database, tRNA gene predictions</td>
<td><a href="http://lowelab.ucsc.edu/GtRNAadb/">http://lowelab.ucsc.edu/GtRNAadb/</a></td>
</tr>
<tr>
<td>UniProt</td>
<td>Protein resource and functional information</td>
<td><a href="http://www.uniprot.org">http://www.uniprot.org</a></td>
</tr>
<tr>
<td>UniProtKB</td>
<td>Curated protein database (UniProtKB/Swiss-Prot and UniProtKB/TrEMBL)</td>
<td><a href="http://www.uniprot.org/help/uniprotkb">http://www.uniprot.org/help/uniprotkb</a></td>
</tr>
<tr>
<td>Gene Ontology (GO)</td>
<td>GO database, annotation of genes, protein and sequences</td>
<td><a href="http://www.geneontology.org/">http://www.geneontology.org/</a></td>
</tr>
<tr>
<td>METACYC</td>
<td>Database for metabolic pathways</td>
<td><a href="http://metacyc.org">http://metacyc.org</a></td>
</tr>
</tbody>
</table>

(Ali et al., 2013)

Table 2.2: Gene Prediction tools, an automatic and manual annotation pipelines, databases and resources. Tools for comparative genomics/proteomics analysis.
2.3.3.1 Gene prediction

Gene identification is the first task after completion of any genome sequencing. If a putative protein encoded by an uncharacterized ORF shows statistically significant similarity to another protein of known function, the ORF in question is a new gene and predicts its likely function (Borodovsky et al., 1994). Even if the homolog of the new protein has not been characterized, useful information is produced in the form of conserved motifs that may be important for protein function. The methods of choice for the initial database screening are those, that translate the query nucleotide sequence in all six reading frames and compare the resulting putative protein sequences to the protein sequence database (Gish and States, 1993).

Sequence analysis methods that distinguish between coding and non-coding regions in DNA on the basis of their different statistical properties are therefore indispensable for gene identification (Fickett and Tung, 1992; Gelfand, 1995). The non-homogenous Markov models using in-phase hexamer statistics (Borodovsky et al., 1994; Krogh et al., 1994), and hidden Markov models (Krogh et al., 1994), have proved particularly effective in bacterial gene prediction. Many of the genes originally predicted by these statistical methods have subsequently proved to be homologous to newly described genes or have been confirmed experimentally, thus supporting the robustness of the prediction methods (Borodovsky et al., 1995).

2.3.3.2 Functional prediction

Most of the sequence similarities are detectable with standard database-searching methods, such as BLASTP (Altschul et al., 1990; Altschul et al., 1994). Additional approaches to similarity analysis, including methods for identifying motifs, produce a significant increase in sensitivity. Proteins with closely related homologs, database screening
with conserved motifs frequently provides additional connections to functionally well-characterized proteins.

### 2.3.3.3 Comparative genomics

Comparative genomics is the study of the differences and similarities in genome structure and organisation in different organisms. Comparative genome sequencing is an important tool in the ongoing effort to exploit conservation in order to annotate and analyse genes and architectural features of genomes. Comparative genomic analysis can be used to define the basic concepts to describe and understand genome evolution (Wolfe and Li, 2003), which is closely linked to gene evolution. The relationship between evolutionary and bioinformatics analyses is evidently reciprocal and synergistic. The intimate relationship between evolution and bioinformatics analysis is nicely illustrated by the fact that one of the first computational analyses of sequences was a phylogenetic analysis, i.e. a study of molecular evolution (Fitch and Margoliash, 1967; Prager and Wilson, 1978). Based on these bioinformatic studies of sequences, many important and intrinsically relevant results for the study of evolution have been obtained. Comparative genomics has revolutionised taxonomy and the understanding of the interplay between phenotype and genotype (Forst and Schulten, 2001). Understanding relationships between the genomes of different species can yield insight into many aspects of evolution, and is especially valuable for the identification of genes and regulatory regions.

### 2.3.3.4 Finishing a sequencing project (Genome finishing)

Genome finishing involves determining the order and orientation of the consensus sequences of contigs obtained from Phrap assemblies of random draft genomic sequences (Lee and Vega, 2004; de la Bastide and McCombie, 2007). This process consists of linking contig ends using information embedded in each sequence file that relates the sequence to the
original cloned insert. Since inserts are sequenced from both ends, a link can be established between these paired ends in different contigs, and thus the contigs can be ordered and orientated. Since genomes may carry numerous copies of insertion sequences, these repeated elements confuse the Phrap assembly program.

It is thus necessary to break these contigs apart at the repeated sequences and individually join the proper flanking regions using paired-end information, or using results of comparisons against a similar genome. Larger repeated elements such as the small subunit ribosomal RNA operon require verification using polymerase chain reaction (PCR) amplification and sequencing. Tandem repeats require manual intervention and typically rely on single nucleotide polymorphisms to be resolved. Filling remaining gaps requires PCR amplification and sequencing. Once the genomes have been closed, low quality regions are addressed by re-sequencing reactions.

For smaller, less complex genomes, the computational scaffolding process may be able to produce complete chromosomes. However, for more complex organisms, the scaffolding process will still result in hundreds or thousands of scaffolds that must be sorted into chromosomes and ordered and oriented. If a closely related genome has already been finished, this may be done computationally. Additionally, depending on the quality of sequence desired, any gaps between scaffolds may need to be filled in using very low-throughput techniques. Most non-model organisms never make it to a high degree of finishing and such genome sequences are known as draft assemblies.
MATERIALS AND METHODS
CHAPTER III
MATERIALS AND METHODS

The present study was undertaken to characterize and annotate whole genome sequence of *Pasteurella multocida* isolated from different species of animals and identification of virulence associated genes.

3.1 ISOLATION AND CONFIRMATION OF *P. multocida* FROM DIFFERENT SPECIES

3.1.1 Collection of samples and isolation of *P. multocida*

All the samples suspected of *Pasteurella multocida* were collected from four different animal species viz. buffalo, cattle, goat and poultry, whereas P52 is a vaccine strain used for vaccination against Haemorrhagic septicaemia. Details of the isolates are given in Table 3.1.

Table 3.1: Details of *Pasteurella multocida* isolates from different animal species

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Species</th>
<th>Isolate designation</th>
<th>Place</th>
<th>Type of tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Buffalo</td>
<td>Anand1_buffalo</td>
<td>Kapadvanj, Kheda district</td>
<td>Lung</td>
</tr>
<tr>
<td>2</td>
<td>Cattle</td>
<td>Anand1_cattle</td>
<td>Gutal, Nadiad, Kheda district</td>
<td>Liver</td>
</tr>
<tr>
<td>3</td>
<td>Goat</td>
<td>Anand1_goat</td>
<td>RBRU farm, Veterinary College, Anand district</td>
<td>Nasal swab</td>
</tr>
<tr>
<td>4</td>
<td>Poultry</td>
<td>Anand1_poultry</td>
<td>Adas, Anand district</td>
<td>Liver</td>
</tr>
<tr>
<td>5</td>
<td>Vaccine Strain P52</td>
<td>P52VAC</td>
<td>Animal Vaccine Institute, Gandhinagar district</td>
<td>Blood</td>
</tr>
</tbody>
</table>

Blood agar (BA) was used as a primary culture medium for preliminary isolation of the *P. multocida* isolates from the suspected samples. A loopful of
triturated tissue sample was first inoculated on BA and incubated at 37°C for 24 hours. Next day, non-haemolytic single colony from BA was transferred to MacConkey agar (MCA) and incubated at 37°C for 24 hours. The isolates which failed to grow on MCA was preliminary presumed to be \textit{P. multocida}. Single non haemolytic colony of these isolates from the primary culture was re-streaked on fresh BA plate and incubated at 37°C for 24 hours to obtain single colony of pure culture of the isolates. The isolates were obtained and stored on blood agar slant in duplicate for further identification by biochemical tests and other studies. All the isolates were stained by Gram’s method and observed under microscope to check the purity of growth.

3.1.2 Biochemical characterization of \textit{P. multocida} isolates

Presumptive \textit{P. multocida} isolates were primarily identified by biochemical tests \textit{viz.}, oxidase, catalase, indole, citrate utilization and nitrate reduction as per the methods described by Barrow and Feltham (1993) with some minor modifications.

3.1.2.1 Oxidase test

Standard oxidase discs (HiMedia, Mumbai) were used to perform the test. A single colony was just touched on the disc with platinum loop/ match stick. Immediate development of blue colour was considered as positive.

3.1.2.2 Catalase test

This test was performed by taking 2-3 drops of three per cent H$_2$O$_2$ on clean grease free glass slide and single colony was mixed with the help of wire loop. Immediate formation of gas bubbles was considered as positive test.

3.1.2.3 Indole test

Few drops of xylene were added to a two day old broth of the isolate in two mL of tryptone water and mixed thoroughly to dissolve indole and about 0.2 mL of
Kovac’s reagent (HiMedia, Mumbai) was added. Pink ring of xylene was considered as positive reaction.

### 3.1.2.4 Citrate utilization

Slant of Simmon’s citrate agar (HiMedia, Mumbai) was inoculated with the culture and incubated at 37°C for four days. Growth with a development of intense blue color of the medium was considered as a positive reaction.

### 3.1.2.5 Nitrate reduction

Added few drops of two days old broth culture in 2 mL peptone water containing 0.1 per cent potassium nitrate and then incubated at 37°C for two days. Presence of nitrate was detected by adding approximately 4-5 drops of sulfanilic acid and 4-5 drops of alpha naphthylamine reagent to nitrate broth culture. Development of a distinct red colour (which may turn to brown rapidly) was considered as positive test.

### 3.1.3 PCR based detection of *P. multocida*

#### 3.1.3.1 DNA Extraction

The genomic DNA of *P. multocida* isolates were extracted according to Wilson (1987) with minor modification. The culture was prepared by inoculating the isolates in BHI (Brain heart infusion) broth and incubated at 37°C for 24 hours in a shaker water bath.

##### 3.1.3.1.1 Reagents used for DNA extraction

- Tris-EDTA (pH 8.0)
  
  10mM Tris-HCL
  
  1mM EDTA

- Sodium dodecyl sulphate (SDS) (10% w/v)

- Proteinase K solution (20mg/mL, w/v)
Materials and Methods...

- Lysozyme (10mg/mL)
- 5M Sodium Chloride
- 7.5M Ammonium acetate
- 10% cetylmethyl ammonium bromide (CTAB)
- Saturated phenol (pH 8.0)
- Chloroform
- Isoamyl alcohol
- Absolute ethanol
- 70% ethanol
- RNAse A
- Nuclease free water

3.1.3.2 Procedure of genomic DNA isolation by Proteinase K-SDS method

- Overnight grown culture cells (10 mL) in BHI broth were centrifuged at 10,000 rpm for 10 minutes.
- Supernatant was discarded and pellet was resuspended in 2 mL Tris-EDTA (pH 8.0).
- To resuspended pellet, 250 μL SDS (10% w/v), 10 μL of proteinase K solution (20 mg/mL; Qiagen, Germany) and 50 μL of lysozyme (20mg/mL; Fermentas, USA) was added, and incubated for 1 hour at 37°C.
- After adding 500 μL of 5M NaCl and 100 μL CTAB, the tube was incubated in water bath for 10 minutes at 65°C and then centrifuged at 10,000 rpm for 10 minutes at 4°C.
- Supernatant was transferred to fresh microfuge tube and equal volume of phenol: chloroform: isoamyl alcohol (25:24:1) was added and
Materials and Methods...

Centrifuged at 10,000 rpm for 10 minutes at 4°C and upper phase was transferred to clean microfuge tube.

- Equal volume of Chloroform: Isoamyl alcohol (24:1) was added, mixed well by inverting, spun for 10 min. at 10,000 rpm and upper aqueous phase was transferred to clean microfuge tube.

- Supernatant was transferred in fresh microfuge tube in which 1/10\textsuperscript{th} volume of 7.5M ammonium acetate and double volume of chilled absolute ethanol was added followed by centrifuging at 12,000 rpm for 15 minutes at 4°C and ethanol was discarded.

- The pellet was washed in 70% ethanol, followed by centrifugation at 11,000 rpm for 5 minutes and finally dried at 65°C for ~10 minutes.

- Pellet was resuspended in 250 µL of nuclease free molecular biology grade water, 2 µL of RNase A was added and incubated at 37°C for 1 hour and then at 75°C for another 15 minutes.

3.1.3.3 Quality checking and quantitation of DNA

Quality and quantity of DNA was calculated by using Nanodrop 1000 spectrophotometry (Thermo Scientific, USA) as well as agarose gel electrophoresis (0.8%). Optical density (OD) at 260 and 280 were taken in spectrophotometer with distilled water as reference.

3.1.3.4 PM-PCR for identification of \textit{P. multocida}

For PM-PCR based identification of \textit{P. multocida} genomic DNA, following primer sequence as described by Townsend \textit{et al.} (1998) and yielding the expected product of 460bp for the subsequent use in GS FLX Titanium (Roche, Switzerland) sequencing experiment.
3.2 LIBRARY PREPARATION OF *P. multocida* DNA

3.2.1 DNA fragmentation (Nebulization)

- Before starting DNA library preparation, quality and quantity of DNA was checked with help of Nanodrop spectrophotometer (Thermo Scientific, USA) and agarose gel electrophoresis.
- Nebulizer unit was assembled and 500ng of DNA was loaded at bottom of nebulizer cup and TE buffer (pH 8.0) was added to make final volume to 100 μL, followed by 500 μL of nebulizing buffer (Roche, Switzerland).
- Nitrogen gas was vented through nebulizing unit at 30 psi for 1 minute, and the sample was collected carefully so to get recovery greater than 300 μL. To this, 2.5 mL of buffer PBI (Qiagen, Germany) was added and swirled properly.
- DNA fragments were purified by using mini elute columns.
- To the column, 750 μL of the sample was added and centrifuged at 10,000 rpm for 15 seconds.
- The flow through was discarded and the above step was repeated thrice in the same column. Finally centrifuged for a minute and the remaining solution was discarded.
- Added 750 μL of PE buffer (Qiagen, Germany), centrifuged at 13,000 rpm for one minute after which, the flow through was discarded.
- The tubes were centrifuged for 15 sec. and then rotated 180° and centrifuged for another 15 seconds.
The column was transferred to a 1.5 mL centrifuge tube and 16 µL of elution buffer (TE buffer) was added.

Centrifuged for 1 minute at 13,000 rpm. The eluant was again passed through the same column and centrifuged.

The eluted sample (approximately 16 µL), was transferred to 200 µL PCR tubes.

3.2.2 Fragment end polishing

End repair mix preparation

Table 3.2: Preparation of end polishing reaction mixture

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL 10X PNK Buffer</td>
<td>2.5 µL</td>
</tr>
<tr>
<td>RL ATP</td>
<td>2.5 µL</td>
</tr>
<tr>
<td>RL dNTP</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>RL T4 polymerase</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>RL PNK</td>
<td>1.0 µL</td>
</tr>
<tr>
<td>RL Taq polymerase</td>
<td>1.0 µL</td>
</tr>
<tr>
<td><strong>Total volume</strong></td>
<td><strong>9.0 µL</strong></td>
</tr>
</tbody>
</table>

To the PCR tubes containing the samples, 9 µL of end polishing reaction mixture was added.

The tubes were vortexed followed by centrifugation at 10,000 rpm for few seconds.

The samples were run in the thermocycler (Applied Biosystems) with the below set parameters.

- 25° C for 20 minutes
- 72° C for 20 minutes
- 4° C on hold.
3.2.3 Bead preparation

✓ The AMPure beads (Backman Coulter, USA) were vortexed for 20 seconds.
✓ Fractionated 125 µL of the beads to a 2 mL centrifuge tube.
✓ Placed the tubes in a magnetic particle concentrator (MPC; Invitrogen, USA) until all the beads pelleted to one side.
✓ The supernatant was then pipetted out carefully.
✓ To the tubes, 73 µL of TE buffer was added and the tubes were vortexed for 5 seconds.
✓ To the beads, 500 µL of sizing solution was added, vortexed for 5 seconds and centrifuged for 2 seconds.
✓ The tubes were then placed on ice until further use.
✓ In a tube, 5 mL of 70% ethanol was prepared by mixing 3.5 mL of 100% ethanol and 1.5 mL of molecular grade water (Milli-Q, Merck Millipore, USA).

3.2.4 Adaptor ligation

✓ To each of the tubes, 1 µL of RL (Rapid library) MIDs adaptors (Roche, Switzerland) was added, after the completion of the end repair programmed cycles.
✓ To the mixture, 1 µL of Ligase was then added, the tubes were vortexed and centrifuged in a micro centrifuge for 2 seconds at 10,000 rpm.
✓ The tubes were incubated at 25 °C for 10 minutes.

3.2.5 Small fragments removal

✓ To the bead solution, 73 µL of the sample was added. Vortexed and spun in the microcentrifuge for 5 seconds at 10,000 rpm.
After incubation, the tubes were placed on the MPC until all the beads were pelleted and carefully discarded the supernatant.

- To the tubes, 100 μL of TE buffer was added and vortexed for 5 seconds.
- To the tubes, 500 μL of sizing solution was added and vortexed for 5 seconds.
- The tubes were then incubated at room temperature for 5 minutes.
- They were then placed on the MPC till all the beads pelleted on the wall and discarded the remaining supernatant.

The above steps were repeated once. The tubes were placed on the MPC and the beads were washed twice with 70% ethanol followed by their complete removal.

- The tubes were then air dried for 2 minutes at room temperature.
- To the tubes, 53 μL of TE buffer was then added, vortexed for 5 seconds and centrifuged for 2 sec. in a microcentrifuge.

- The tubes were placed on the MPC, the beads were pelleted and 50 μL of supernatant which contained the library was transferred to a new centrifuge tube.

3.2.6 Library quantitation

3.2.6.1 Preparation of standard for TBS 380 fluorometer

- For generation of standard curve, 8 tubes were labeled (1-8).
- Tube 1: 2.5 X 10⁹ molecules/μL solution of RL standard (Roche, Switzerland) was prepared by mixing 90 μL of RL standard with 90 μL of TE Buffer.
- The remaining 7 tubes (2-8) were filled with 60 μL of TE buffer.
- From tube 1 to tube 2, 120 μL of solution was transferred.
Materials and Methods...

✓ The tube was vortexed for 5 seconds and spun in a minicentrifuge for 2 seconds.
✓ From tube 2 to tube 3, 120 μL of solution was transferred.
✓ The tube was vortexed for 5 seconds and spun in a minicentrifuge for 2 seconds.
✓ The serial dilution was proceeded as above for the remaining 5 tubes.

3.2.6.2 Sample library quantitation using TBS 380 fluorometer

✓ To 8 cuvettes, 50 μL of 8 dilutions of the RL standard were transferred.
✓ As the blank, 50 μL of TE buffer was taken.
✓ The TBS 380 fluorometer (Promega, USA) was set on the blue channel with the blue cuvette holder insert.
✓ The standard value was set to 250.
✓ The fluorometer was calibrated with the blank and 2.5 x 10^9 molecules/μL solution RL standard.
✓ The relative fluorescence units (RFU) was read and recorded for each dilution.
✓ To a cuvette, 50 μL of the sample library was transferred. The fluorescence was read and recorded.
✓ The sample library was transferred back to its tube with 20 μL pipette.

3.2.6.3 Generation of rapid library (RL) standard curve and calculation of sample concentration

✓ To generate the standard curve of the fluorescence readings and to calculate the library sample concentration, the rapid library quantitation calculator was used (http://454.com/my454/tools-downloads/rapid-library calc.asp).
3.2.6.4 Library quality assessment

✓ To assess the quality of library, 1 μL of aliquot of the DNA library was run on a Bioanalyzer High Sensitivity DNA chip (Agilent Technologies, USA).

✓ The quality of the DNA library was assessed for the below characteristics.

Table 3.3: Quality assessment of 454 library

<table>
<thead>
<tr>
<th>Library characteristics</th>
<th>Expected results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average fragment length</td>
<td>Between 600 bp to 900 bp</td>
</tr>
<tr>
<td>Lower size cut off</td>
<td>&lt;10% below 350 bp</td>
</tr>
</tbody>
</table>

3.3 EMULSION PCR

3.3.1 Emulsion based clonal amplification

3.3.1.1 Preparation of the live and mock amplification mixture

✓ All the reagents used for emPCR other than enzymes were thawed, vortexed for 5 seconds and centrifuged (including enzymes) for 10 seconds on bench top minicentrifuge.

✓ The Live amplification mix for small volume (SV) as well as for large volume (LV) were prepared as per Table 3.4 and vortexed for 5 seconds and stored at +2 to +8°C till used.
The mock amplification mix was prepared by diluting 2 mL of 5X mock amplification mix with 8 mL of molecular biology grade water in a 15 mL Falcon tube, vortexed for 10 seconds and then stored at +2 to +8°C till used.

### 3.3.2 DNA library capture

- Small volume emPCR was first done for getting best ratio of molecule/beads to set large volume emPCR.
- Capture bead wash buffer TW (Roche, Switzerland) was diluted 10 times to get 1X capture bead wash buffer TW (10 mL) with molecular biology grade water.
- Single tube of DNA capture beads was vortexted, and centrifuged for 10 seconds, rotating tube 180° and again centrifuged for 10 seconds, followed by discarding supernatant.
The beads were washed twice with 1 mL of 1X capture bead buffer TW and supernatant was discarded by spinning tubes as done in above step.

Washed DNA capture bead pellet was resuspended in 640 μL of 1X capture bead wash buffer and then evenly distributed in 1.7 mL tubes.

The beads were pelleted and correct amount of DNA library (after heat denature at 95°C for 2 minutes) was added to each tube and vortexed for 5 seconds. The amount of DNA library to be added was calculated on basis of required DNA molecules per bead using formula;

\[
\text{μL of library per tube} = \frac{\text{Molecules per bead desired} \times \text{Number of beads per tube}}{\text{Library concentration (molecules/μL)}}
\]

Number of beads per tube = 2.4 X 10^6

For small volume emPCR, sample dsDNA library after fluorometer analysis, was diluted in TE buffer to obtain 1 X 10^7 molecules/μL. From 1 X 10^7 molecules/μL of DNA library, 20 μL of dsDNA library was further diluted into 30 μL of TE buffer, to prepare a 4 X 10^6 molecules/μL. Different amount of dsDNA library and beads were added as per Table 3.5 and vortexed for 5 seconds.

Table 3.5: Amount of library to be used for various molecules/bead ratio

<table>
<thead>
<tr>
<th>Tube no.</th>
<th>Amount of lib. (μL)</th>
<th>Molecules/bead</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube 1</td>
<td>1.2</td>
<td>2.0</td>
</tr>
<tr>
<td>Tube 2</td>
<td>2.4</td>
<td>4.0</td>
</tr>
<tr>
<td>Tube 3</td>
<td>4.8</td>
<td>8.0</td>
</tr>
<tr>
<td>Tube 4</td>
<td>9.6</td>
<td>16.0</td>
</tr>
</tbody>
</table>
Materials and Methods...

✓ Finally, large volume emPCR was done by taking appropriate ratio of DNA library and capture bead from recovery of small volume emPCR.

✓ Capture bead wash buffer TW was diluted 10 times to get 1X capture bead wash buffer TW (10 mL) with molecular biology grade water.

✓ The two tubes of DNA capture beads were vortexed and pelleted by centrifuging twice for 10 seconds by rotating microfuge tube to 180°.

✓ The supernatant was discarded and beads were washed twice with 1 mL 1X bead wash buffer, vortexed for 5 seconds and again pelleted.

✓ DNA library was thawed and correct aliquot was added to each tube of DNA capture bead.

✓ The amount of library used was calculated by following equation:

\[
\text{\( \mu l \) of library per tube} = \frac{\text{Molecules per bead desired} \times \text{Number of beads per tube}}{\text{Library concentration (molecules/\( \mu l \))}}
\]

Number of beads per tube = 35 X 10^6

3.3.3 Emulsification and amplification for small volume

✓ Emulsion oil was pre-mixed on the tissue lyser (Qiagen, Germany) at 25 Hz for 2 minutes.

✓ In each tube of oil, 290 \( \mu L \) of 1x mock amplification was added and inverted 2-3 times.

✓ Content in the tube was mixed at 25 Hz for 5 minutes.

✓ The prepared DNA library (~255 \( \mu L \)) was added to each tube and mixed in tissue lyser at 15 Hz for 5 minutes.

✓ After emulsification, 50 \( \mu L \) of the content was dispensed in 96 well PCR plates (Thermo Fisher Scientific, USA).

✓ Plates were then placed in a thermocycler (Applied Biosystem, USA) with amplification program as follow:
3.3.4 Emulsification and amplification for large volume

- Tubes containing emulsion oil were mounted in tissue lyser and mixed at 28 Hz for 2 minutes.
- To these, 5 mL of 1X Mock amplification mix was added in each cup of oil and again mixed at 28 Hz for 5 minutes in tissue lyser.
- Captured library whole content (~3.75 mL) was added and mixed at 12 Hz for 5 minutes.
- Emulsified emPCR amplification mix was dispensed in 4 plates each with 96 wells (50 μL per well) and sealed.
- Plates were then placed in a thermocycler with amplification program as follows:

<table>
<thead>
<tr>
<th>Number of cycles</th>
<th>Stage 1</th>
<th>Stage 2</th>
<th>Stage 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle parameters</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>1</td>
<td>50</td>
<td>-</td>
</tr>
<tr>
<td>Stage 2</td>
<td>4 min at 94°C</td>
<td>30 sec. at 94°C, 4.5 min at 58°C, 30 sec. at 68°C</td>
<td>10°C until used</td>
</tr>
<tr>
<td>Stage 3</td>
<td>-</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.3.5 Bead recovery (Vacuum assisted emulsion breaking)

✓ Added 187.5 mL of molecular biology grade water (Milli-Q; Millipore, USA) to 62.5 mL of 4X enhancing fluid TW (Roche, Switzerland) stock to make 1X concentrate and placed on ice.

✓ Added 72 mL of molecular biology grade water to 8 mL of 10X annealing buffer TW (Roche, Switzerland) stock to make 1X concentrate and kept on ice.

✓ 50 mL conical tube was fixed to each of the two lids and the connector located at one end of the tubing was inserted into the top opening of the 8 pronged transpette.

✓ The other end of the tubing was connected to a vacuum source.

✓ Vacuum was turned on and emulsions were aspirated from eight wells at a time, using a slow circular motion of the transpette tips at the bottom of the wells.

✓ The amplified DNA beads were collected in two 50 mL tubes. The plate was rinsed twice with 100 µL of isopropanol (Merck, USA) per well.

3.3.6 Bead washes and recovery for large volume emulsions (LVE)

✓ The contents of two (or more) 50 mL collection tubes were mixed by transferring their contents (in pairs) back and forth four times or until the bead suspension were of about the same density.

✓ Isopropanol was added to final volume of 40 mL in each tube and vortexed.

✓ The beads were pelleted in a centrifuge at 930g for 5 min (2,000 rpm).

✓ The supernatant was carefully poured into a waste container and the bead pellets were rinsed twice with 35 mL of isopropanol by centrifugation as mentioned above.
The bead pellets were rinsed once again with 35 mL of 1X enhancing fluid TW.

The DNA bead suspension was then transferred to two 1.7 mL tubes for each emulsion cup processed.

Each of the 50 mL collection tubes were then rinsed with 600 µL of 1X enhancing fluid TW and added this rinse to the 1.7 mL tubes.

Again 1.7 mL tubes were centrifuged in a bench top minifuge.

The supernatant was removed and rinsed twice with 1 mL of 1X enhancing fluid TW.

The bead pellets were then resuspended in a final 1 mL of 1X enhancing fluid TW.

3.3.7 DNA library bead enrichment

Θ Preparation for Indirect Enrichment

At this point, the beads carry double-stranded clonally amplified DNA products, and were suspended in 1X enhancing fluid TW.

A stock was prepared of melt solution by mixing 125 µL of NaOH (10N) in 9.875 mL of molecular biology grade water.

The collected DNA beads were pelleted using bench top minifuge.

Supernatant was discarded and the ssDNA beads were prepared by melting the dsDNA amplification products (only one strand is attached to the beads) using one mL of melt solution per tube, vortexed and incubated for 2 minutes.

The supernatant was removed and discarded.

The beads were washed twice with 1 mL of 1X annealing buffer TW per tube each time.
Materials and Methods...

- 45 μL of 1X annealing buffer TW and 25 μL of enrichment primer was added per tube and vortexed to mix completely.
- The enrichment primers were annealed to the bead-bound ssDNA by placing the tubes in a heat block at 65°C for 5 min and then promptly cooling on ice for 2 min.
- Added 800 μL of 1X enhancing fluid TW and vortexed.
- Centrifuged the beads, pelleted as above, and the supernatant was removed.
- Beads were then washed twice more by centrifugation, with 1 mL of 1X enhancing fluid TW per tube.
- Supernatant was removed, each bead pellet was resuspended in 800 μL of 1X enhancing fluid TW and vortexed.

3.3.8 Preparation of enrichment beads

- The tube of enrichment beads was placed in a MPC, pelleting the paramagnetic enrichment beads against the side of the tube.
- The supernatant was then removed and discarded.
- The pelleted beads were washed twice with 1 mL of 1X enhancing fluid TW, using the MPC.
- At the end after removing the supernatant, the tube was removed from the MPC and 320 μL of 1X enhancing fluid TW was added.

3.3.9 Enrichment of the DNA carrying beads

- To each tube of amplified DNA beads, 80 μL of washed enrichment beads were added and vortexed to mix completely.
- Then rotated on a LabQuake tube roller (Thermolyne) at ambient temperature (+15°C to +25°C) for 5 minutes.
The tubes were placed in the MPC and allowed to stand for 3-5 minutes to pellet the paramagnetic enrichment beads against the side of the tubes.

The supernatant was carefully removed and discarded from each tube using 1000 μL pipette.

The beads were then washed with 1 mL of 1X enhancing fluid TW per tube until there were no visible beads remaining in the supernatant (usually 6-10 washes for LVE)

Removed the tubes from the MPC and bead pellets were re-suspended in 700 μL of melt solution.

Then vortexed for 5 seconds and placed back into the MPC to pellet the enriched beads.

Transferred the supernatants containing enriched DNA beads to a separate 1.7 mL microfuge tube.

Pelleted enriched DNA beads by centrifugation and the bead pellets were washed three times with 1 mL of 1X annealing buffer TW (by centrifugation) to completely neutralize the melt solution.

Each of the final bead pellets were re-suspended in 200 μL of 1X annealing buffer TW.

3.3.10 Sequencing primer annealing

Fifty microliters of sequencing primer was added and vortexed.

Tubes were then placed in a heat block at 65°C for 5 min, and then kept promptly on ice for 2 minutes.

Added 800 μL of 1X annealing buffer TW to the tube, pelleted and the supernatant was removed.

Bead pellets were then washed twice with 1 mL of 1X annealing buffer TW.
Materials and Methods...

✓ Resuspended each of the final bead pellets by adding 1 mL of annealing buffer TW.

✓ Beads were then stored (immobilized, clonally amplified DNA library) at +2°C to +8°C.

3.4 SEQUENCING EXPERIMENT AND DATA PROCESSING

3.4.1 Sample preparation procedure

3.4.1.1 Preparation of bead buffer 2 (BB2)

✓ 1.2 mL of titanium supplement CB (Roche, Switzerland) was added to the 200 mL bottle of pre-chilled titanium bead buffer (Roche, Switzerland) to prepare BB2.

✓ Thirty four microliters of apyrase solution was added to the bottle of titanium bead buffer plus titanium supplement CB, mixed by inversion.

✓ BB2 was poured into the tray to completely submerge the PTP (Pico titre plate; Roche, Switzerland) device.

✓ The PTP device was left submerged in BB2 at room temperature on the lab bench for 10 minutes.

3.4.1.2 Preparation of packing beads

✓ Packing beads were washed three times with BB2 using one microliter of BB2 to the tube of packing beads, vortexed and centrifuged at 10,000 rpm for 5 minutes.

✓ Supernatant was discarded and beads were re-suspended by adding 550 µL of BB2.
3.4.1.3 Preparation of DNA beads (Sample and control)

✓ Sample DNA beads were spiked with control DNA beads included in the GS sequencing kit (Roche, Switzerland) which serves as internal control for sequencing reaction.

✓ DNA library was vortexed and transferred 30 (x4) μL to a 1.7 mL tube. Appropriate volume of control DNA bead 10 (x4) μL was added to each DNA library bead tube.

✓ The beads were centrifuged for 1 minute at 10,000 rpm twice. DNA bead incubation mix (DBIM; Roche, Switzerland) was prepared side by side in 15 mL tube and vortexed.

<table>
<thead>
<tr>
<th>Loading region size</th>
<th>All size</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTP size</td>
<td>70X75 mm</td>
</tr>
<tr>
<td>BB2 (μL)</td>
<td>1570</td>
</tr>
<tr>
<td>Polymerase cofactor (μL)</td>
<td>150</td>
</tr>
<tr>
<td>DNA polymerase (μL)</td>
<td>300</td>
</tr>
<tr>
<td><strong>Total volume (μL)</strong></td>
<td><strong>2020</strong></td>
</tr>
</tbody>
</table>

✓ The appropriate volume of DBIM (320(x4) μL) was transferred to DNA beads, vortexed and placed on LabQuak tube roller for 15 minutes at room temperature.

✓ The DNA beads were combined with the packing beads to make bead layer 2.

3.4.1.4 Preparation of enzyme and PPIase beads (Bead layers 1, 3 and 4)

✓ Vortexed and then pelleted the enzyme beads and the PPIase beads using a magnetic particle concentrator (MPC) as follows.
The two tubes of enzyme beads and one tube of PPIase beads were placed on the MPC till the pellet was formed. Supernatant was discarded carefully with pipettor and removed the tubes from MPC.

Washed the enzyme and PPIase beads three times using 1 mL of BB2 to each tube of enzyme beads and 500 μL to each tube of PPIase beads and vortexed till the beads were resuspended completely.

The tubes were then placed on MPC till the pellet was formed.

Supernatant was discarded with pipettor and the tubes were removed from MPC and original volume of BB2 was added to each tube.

**Table 3.9: Types of beads**

<table>
<thead>
<tr>
<th>Reagent label</th>
<th>No. of tubes</th>
<th>Volume/tube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme beads</td>
<td>2</td>
<td>1000 μL</td>
</tr>
<tr>
<td>PPIase beads</td>
<td>1</td>
<td>500 μL</td>
</tr>
</tbody>
</table>

The beads for layers 1, 3 and 4 were prepared in three separately labelled tubes of an appropriate size. This was done by diluting the enzyme beads (layer 1: Enzyme beads pre-layer; and layer 3: Enzyme beads post-layer) and the PPIase beads (layer 4) in BB2 as follow.

**Table 3.10: Preparation of beads for layers 1, 3 and 4**

<table>
<thead>
<tr>
<th>Bead layer</th>
<th>BB2 (μL)</th>
<th>Enzyme beads (μL)</th>
<th>PPIase beads (μL)</th>
<th>Total volume (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layer 1</td>
<td>3250</td>
<td>550</td>
<td>---</td>
<td>3800</td>
</tr>
<tr>
<td>Layer 3</td>
<td>2500</td>
<td>1300</td>
<td>---</td>
<td>3800</td>
</tr>
<tr>
<td>Layer 4</td>
<td>3340</td>
<td>---</td>
<td>460</td>
<td>3800</td>
</tr>
</tbody>
</table>
3.4.1.5 Combining DNA and Packing Beads

✓ On completion of incubation of DNA beads with DBIM (DNA bead incubation mix), packing beads and BB2 were added to DNA beads as per Table 3.11.

Table 3.11: Packing beads and BB2 addition to DNA beads

<table>
<thead>
<tr>
<th>Loading region size</th>
<th>Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTP size</td>
<td>70X75 mm</td>
</tr>
<tr>
<td>DNA Beads μL</td>
<td>350 (x4)</td>
</tr>
<tr>
<td>Packing beads μL</td>
<td>100</td>
</tr>
<tr>
<td>BB2 μL</td>
<td>210</td>
</tr>
<tr>
<td>Total beads μL</td>
<td>660</td>
</tr>
</tbody>
</table>

✓ The DNA and packing bead mixture was then incubated on LabQuake tube roller for 5 minutes, and unused packing beads were discarded.

✓ PTP device was removed, wiped and placed on BDD (Bead deposition device).

✓ Bead loading gasket was placed on BDD base by laying it on PTP and assembled.

3.4.2 Picotitreplate (PTP) device preparation

3.4.2.1 Bead layer 1: Enzyme beads pre-layer

✓ The bead suspension for layer 1 was vortexed for 5 seconds and 660 μL was carefully drawn for uniform bead distribution throughout region.

✓ The assembled BDD along with counterweight were centrifuged for 5 minutes at 1620g, seals were removed and BDD was discarded using pipettor.
3.4.2.2 Bead layer 2: DNA and packing beads

✓ Tube of beads of layer 2 was centrifuged and mixed.
✓ Six hundred sixty μL was dispensed over the first layer through loading port and sealed.
✓ The assembled BDD along with counterweight were centrifuged for 10 minutes at 1620g, seals were removed and BDD was discarded.

3.4.2.3 Bead layer 3: Enzyme beads post-layer

✓ Tube of enzyme beads of layer 3 was centrifuged and mixed.
✓ Six hundred sixty μL was dispensed over the first layer through loading port and sealed.
✓ The assembled BDD along with counterweight were centrifuged for 5 minutes at 1620g, seals were removed and BDD was discarded.

3.4.2.4 Bead layer 4: PPiase beads

✓ Tube of PPiase beads was centrifuged briefly and mixed.
✓ Six hundred sixty μL was dispensed to over first layer through loading port and sealed.
✓ The assembled BDD along with counterweight were centrifuged for 5 minutes at 1620g, seals were removed and BDD was discarded.

3.4.3 Sequencing run

3.4.3.1 Preparation and loading the sequencing reagents cassette

✓ The titanium supplement CB (Roche, Switzerland) was added to each bottle of titanium buffer CB as per Table 3.12.
Table 3.12: Addition of titanium supplement CB

<table>
<thead>
<tr>
<th>Reagent label</th>
<th>No. of bottles</th>
<th>Volume of titanium supplement CB added per bottle</th>
<th>Volume of DTT added per bottle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titanium buffer CB</td>
<td>4</td>
<td>6.6 mL</td>
<td>1000 µL</td>
</tr>
</tbody>
</table>

✓ Forty five µL of inhibitor TW was transferred from the sequencing reagent tray to 1.7 mL tube.

✓ Five µL of the concentrated PPiase reagent was added to 45 µL aliquot of inhibitor TW and vortexed.

✓ 13.2 µL of the diluted PPiase reagent was added to inhibitor TW tube.

✓ 260 µL of apyrase reagent to the tube of the buffer for apyrase reagent.

✓ 3000 µL of dATP reagent to the tube of buffer for dATP.

✓ The entire sequencing reagent tray was inverted several times to uniformly mix the contents of all tubes and the sequencing reagent tray loaded into the instrument.

3.4.3.2 PTP cartridge loading and sequencing run

✓ PTP was carefully lifted and inserted into cartilage from BDD after deposition of forth layer, discarding supernatant and uncasing top of BDD.

✓ Started the sequencing run using various parameters for sequencing run PTP id, Name of sequencing run, PTP type (4 Region), Number of cycles (200) for nucleotide flow and sequencing run type was full processing.

3.4.4 Data transfer and signal processing

✓ The data generated was transferred to GS FLX cluster with help of file transfer protocol (FTP) and signals present in form of images were
Materials and Methods...

processed, giving reads using "GS Run Browser" software (Roche, Switzerland).

✓ The sequencing depth was calculated using following formula:

\[
\text{Sequencing depth (coverage)} = \frac{\text{Total number of bases in reads}}{\text{Total number of bases in reference}}
\]

3.5 BIOINFORMATICS ANALYSIS

The reads generated using 454 sequencing run were further analysed using various bioinformatics tools. Preliminary analysis was done using two different softwares viz., GS Reference Mapper (Roche, Switzerland) for mapping the genome of *Pasteurella multocida* isolates and another was GS De Novo Assembler (Roche, Switzerland), used for assembling reads which were not assigned to the reference genome. Remaining reads which were not mapped and assembled were singleton reads and used for the BLASTX analysis. The detailed procedure for above software is given below.

3.5.1 GS Reference Mapper

The GS Reference Mapper (Roche, Switzerland) aligns sequencing reads against a reference sequence consisting of one or more sequences, with or without associated annotations. The GS Reference Mapper software is an interactive application used to create mapping projects, add or remove reads from the project, specify reference sequences, annotations and other project parameters, run the mapping algorithms on the project data, and view the output produced by the mapping computations. The application can be accessed via a Graphical User Interface (GUI) or from a command line interface (CLI).

Mapping generates consensus sequences of the reads that align against the reference and also computes statistics for variations found in the reads, relative to the
reference. Data were output to a variety of file formats, including FASTA, ACE, BAM or consed files.

When the mapping algorithms run, the software performs the following operations:

- For each read, search for its best alignment to the reference sequence(s) (a read may align to multiple positions in the reference); this is done in ‘nucleotide’ space.
- Perform multiple alignments for the reads that align contiguously to the reference in order to form “contigs.” From the contigs’ multiple alignments, consensus base call sequences were produced using the signals of the reads in the multiple alignments (performed in ‘flowspace’).
- Identify subsets of the reads that vary relative to the reference to form lists of putative variations (nucleotide differences and structural variants). For each putative variation, the reads supporting the variation will be in the subset.
- Evaluate these lists of putative variations to identify High-Confidence nucleotide differences (HCDiffs), structural variations (HCStructVars) and larger-scale structural rearrangements (HCStructRearrangements).

- Output the following information:
  ✓ Contig consensus sequence(s) and associated quality values
  ✓ Alignments of the reads to the reference
  ✓ Position-by-position metrics of the depth and consensus accuracy (quality values) for each position in the aligned reference
  ✓ The positions and alignments of identified differences.

For performing mapping of our five isolates of Pasteurella multocida, we used Pm70, isolated from poultry as a reference sequence available for the P. multocida.
The mapping project was created using reads that were available with 454 pyrosequencing approach. Multiplex Identifiers (MIDs) tab was selected to scan for the presence of MIDs, and the reads containing the selected MIDs were made eligible for mapping in the project. The input parameters kept for mapping were kept with (Seed step- 12, Seed length- 16, Seed count- 1, Hit-per-seed limit- 70, Minimum overlap length- 40, Minimum overlap identity- 90, Alignment identity score- 2, Alignment difference score- -3, Repeat score threshold- 12) for all Pasteurella isolates.

3.5.2 GS De Novo Assembler

The GS De Novo Assembler (Roche, Switzerland) application constructs De Novo assemblies of the reads from one or more sequencing runs, using the “read flowgrams” (SFF files) as input. The GS De Novo Assembler software, an interactive application used to create assembly projects, add or remove reads from the project, specify project parameters, run the assembly algorithms on the project data, and view the output produced by the assembly computations. The application can be accessed via a graphical user interface (GUI) or from a command line interface (CLI).

Input data was combined from one or several regions of one or several runs of interest. Assembly generates a consensus sequence of the sample DNA library, output as one or more contiguous sequences (contigs) in FASTA, ACE or consed files.

During the assembly process, the software:

✓ Identifies pairwise overlaps between reads,

✓ Constructs multiple alignments of overlapping reads and divides or introduces breaks into the multiple alignments in regions where consistent differences were found between different sets of reads. This step results in a preliminary set of “contigs” that represents the assembled reads.
Materials and Methods...

- Attempts to resolve branching structures between contigs,
- Generates consensus base calls of the contigs by using quality and flow signal information for each nucleotide flow included in the contigs multiple alignments,
- Outputs the contig consensus sequences and corresponding quality scores, along with an ACE file of the multiple alignments and assembly metrics files.

Read overlaps and multiple alignments were made in “nucleotide” space while the consensus base calling and quality value determination for contigs were performed in “flowspace”. Work in flowspace allows the quality-weighted averaging of processed flow signals (a continuous variable) at each nucleotide flow of the sequencing run(s) and allows the use of information from the “negative flows”, i.e. flows where no nucleotide incorporation is detected. The use of flowspace in determining the properties of the consensus sequence results in an improved accuracy for the final base calls.

The assembler produces contigs from the multiple alignments of overlapping read sequences. The GUI provides tools that allow the contig consensus sequences, the multiple alignments of the reads that form the contig, and the flowgrams of these reads to be viewed interactively.

The new assembly project was created in GS De Novo Assembler and when a project opens, the overview tab was displayed. The read data file (.sff) was added before an assembly computation was performed. Use Multiplex Filtering” checkbox was selected which allows to screen multiple libraries of Pasteurella multocida prepared using distinct MID tags and sequenced together, on the same PTP device. The input parameter for assembly were kept (Seed step-12, Seed length- 16, Seed
count- 1, Minimum overlap length- 40, Minimum overlap identity- 90, Alignment identity score- 2, Alignment difference score- -3) for all *P. multocida* isolates.

### 3.5.3 RAST annotation server

RAST server was developed by Aziz *et al.* (2008), designed to rapidly call and annotate the genes of a complete or essentially complete prokaryotic genome. RAST, Rapid Annotations based on Subsystem Technology, uses a "Highest Confidence First" assignment propagation strategy based on manually curated subsystems and subsystem-based protein families that automatically guarantees a high degree of assignment consistency. For analysis of all our five *Pasteurella* isolates, contigs generated after reference mapping with Pm70 were uploaded to RAST server (http://rast.nmpdr.org/).

RAST used existing tools built by other research teams to first identify both the tRNA and rRNA encoding genes. For the tRNA genes, it uses tRNAscan-SE (Lowe and Eddy, 1997) and to identify the rRNA encoding genes, uses a tool "search_for_rnas" (McNeil *et al.*, 2007). Then, the server will not consider retaining any protein-encoding gene that significantly overlaps any of these regions.

Once the tRNA and rRNA gene-encoding regions were removed from consideration, an initial call using GLIMMER2 was made. Using this small set of representatives, it searches the protein-encoding genes from the new genome for occurrences of these Fellowship of Interpretation of Genome families (FIGfams). This was a very rapid step, since only the new genome was being searched, and it was being searched using a small set of representative protein sequences. The outcome of this initial scan was a small set (normally, 8-15 genes) that can be used to estimate the closest phylogenetic neighbours of the newly-sequenced genome. The putative genes
that remain can be used to search against the entire collection of FIGfams. This was done by blasting against a representative set of sequences from the FIGfams to determine potential families that need to be checked, and then checking against each family.

The final assignments of functions were made to the remaining putative genes. If similarities were computed in the preceding step, these similarities can be accessed and functions can be asserted. Once assignments of function was made, an initial metabolic reconstruction was formed.

3.5.4 NCBI Prokaryotic Genome Automatic Annotation Pipeline (PGAAP)

NCBI Prokaryotic Genome Automatic Annotation Pipeline was designed to annotate bacterial and archaeal genomes (chromosomes and plasmids). Genome annotation, a multi-level process that includes prediction of protein-coding genes, as well as other functional genome units such as structural RNAs (5S, 16S and 23S), tRNAs, small RNAs, pseudogenes, control regions, direct and inverted repeats, insertion sequences, transposons and other mobile elements (http://www.ncbi.nlm.nih.gov/genome/annotation_prok/).

PGAAP developed by NCBI is an automatic annotation pipeline that combines ab initio gene prediction algorithms with homology based methods. The pipeline is capable of annotating both complete genomes and draft WGS genomes consisting of multiple contigs. Since our genome sequenced were not in a single scaffold, the genome assembly is in multiple pieces and were submitted using WGS submission portal (https://submit.ncbi.nlm.nih.gov/subs/wgs1/) after removing the sequences which were less than 200bp (using in house script).
3.5.4.1 Gene finding using GeneMarkS+

A gene prediction program GeneMarkS+, integrates information about protein alignments, frameshifted genes, non-coding RNA, and DNA statistical patterns typical for protein-coding and non-coding regions into gene predictions. GeneMarkS+ is an extension of the GeneMarkS ab initio gene finding program (Besemer et al., 2001).

Detecting frameshifts is a critical component of resolving ambiguities in automated annotation, and provides important feedback in assessing the quality of an assembly. Frameshift detection in the pipeline was implemented as a two-step process. Proteins from the target set were aligned to the genome with ProSplign (a global alignment algorithm) that detects alignments with frameshifts. These regions were passed to GeneMarkS+ and reported in the final output as pseudogenes. In a second step, newly predicted GeneMarkS+ genes were evaluated for potential frameshifts. All the proteins were aligned to the search set used for protein identification and naming. Several categories of potential frameshift were handled currently, including a set of tandem models collocated within a threshold of 200 nucleotides and aligned to the same search protein as well as singleton partial alignments of sufficient quality, were identified as a potential frameshifted region. All candidate search proteins were then aligned to the region with ProSplign and evaluated for frameshifts. If at least four candidate search proteins align in a given region with a frameshift, the original models were replaced with a new gene feature with a pseudo qualifier covering the maximal extent of aligned frameshifted protein.
Materials and Methods...

Figure 3.1: Workflow for the GeneMarkS+ used by NCBI database

Figure 3.2: Overview of gene prediction pipeline
3.5.4.2 Annotation results of PGAAP

The annotation pipeline produces ASN.1 files (*.sqn) ready for GenBank submission. It can also convert ASN.1 to traditional GenBank flat file format for manual review. The summary report is generated as a part of the output file and includes the total number of predictions by each feature type.

3.5.5 tRNA prediction

For tRNA prediction tRNAscan-SE, available at http://lowelab.ucsc.edu/tRNA scan-SE/ was used (Schattner et al., 2005). Raw sequences generated after sequencing were uploaded to tRNAscan-SE 1.21 server for the tRNA prediction.

3.5.6 rRNA prediction

For rRNA prediction, RNAmmer tool was used, available at http://www.cbs.dtu.dk/services/RNAmmer (Lagesen et al., 2007). All the contigs were uploaded to RNAmmer 1.2 server in FASTA format.
RESULTS AND DISCUSSIONS
Bacteria are the most predominant form of free-living life on earth and are ubiquitous. Many of them are causing diseases and hence, understanding their evolution, adaptation, and virulence functions helps in developing intervention procedures involving modern biotechnological tools for our benefit. This also leads to improve our understanding of how the host and bacterial interaction occurs and the mechanisms of overcoming the host defence mechanisms to establish an infection. Bacteria are also excellent systems for studying adaptation, both at the level of genome content and at the level of controlling gene expression and protein activity in response to changing environmental and host conditions.

Genome sequence analysis has enabled the development of various clinico-microbiological tools for pathogen detection, identification or genotyping by identification of sequence fragments specific at distinct taxonomic levels (genus, species, strain, clone), for the detection of genes associated with antibiotic resistance or virulence and for the identification of deficient metabolisms to aid development of optimized culture media. However, whole-genome sequencing (WGS), by giving access to the full genetic repertoire of an isolate, has demonstrated an undisputed discriminatory power for deciphering the whole disease process. In addition to identifying bacteria at various taxonomic levels, WGS also offers an opportunity to detect various genetic markers, such as virulence factors or antibiotic resistance-associated genes. Identifying and characterizing the virulence factors of pathogens are crucial for understanding the pathogenesis of the diseases that they cause and for developing dedicated molecular tools to detect specific virulence markers. Genomics has played an important role in the identification of virulence factors in bacteria.
Three main strategies are used to identify virulence factor encoding genes in
genomes: first, comparison of genomes from strains or species exhibiting diverse
degrees of virulence; second, identification of laterally transferred genomic islands,
assuming that virulence genes are often acquired by this mechanism and third,
comparing the genome against databases of known virulence markers.

Diseases caused by many members of Pasteurellaceae family are very harmful,
aacute and economic loss to the farmers is very high. Despite vaccination to the
animals particularly against Haemorrhagic Septiceamin (HS), several outbreaks occur
in Gujarat as well as in other parts of India on regular basis. Whole genome
sequencing is a recent advanced approach for understanding of genetic makeup of an
organism as well for identification of virulence factors responsible for the disease
process in host. The first whole genome sequence of P. multocida isolated from
poultry was done in 2001 by May et al. (2001). After that, no other attempt for
complete genome sequencing of P. multocida was apparently carried out during the
subsequent decade. Recently, in last 2-3 years, whole genome sequencing of pig,
bovine and poultry (virulent) isolates were carried out (Liu et al., 2012; Abrahante
et al., 2013; Lainson et al., 2013a; Lainson et al., 2013b). None of these whole genome
sequencing projects involved Indian isolates of P. multocida till date, except for Vaid
et al. (2014) and Abrahante et al. (2014) at Minnesota, USA (for the buffalo isolates).
This might be the first study for the whole genome sequencing of P. multocida from
different species of animals to desipher more insight into the virulence factors
responsible for various important diseases in different animal species and poultry.

The study was conducted to sequence whole genome of Pasteurella multocida
of different species, to annotate their genome and finding the virulence associated
genes which were either reported earlier or previously not identified and/or
Results and Discussion...

characterized. These newly identified genes or the protein products may directly or indirectly be involved in virulence of the bacteria and thereby the disease process.

Whole genome sequencing of four *Pasteurella multocida* field isolates viz. poultry, goat, buffalo and cattle as well as P52 (vaccine strain) was carried out at Department of Animal Biotechnology, College of Veterinary Science and Animal Husbandry, Anand Agricultural University, Anand. All the isolates of *P. multocida* were identified by cultural and biochemical characters and were subsequently confirmed by PM-PCR. All the isolates were subsequently processed for the whole genome sequencing using pyrosequencing based 454 GS-FLX Titanium (Roche, Switzerland).

4.1 ISOLATION AND CHARACTERIZATION OF *P. multocida* ISOLATES

For the isolation of *P. multocida*, blood agar (BA) was used as a primary culture medium for preliminary isolation. The samples were initially inoculated on BA plates and incubated at 37°C for 24 hours. The non-haemolytic moderate sized, round, greyish, smooth or mucoid colonies which failed to grow on MacConkey agar (MCA) were primarily presumed to be of *P. multocida*. On the basis of Gram’s staining, the isolates were found to be Gram-negative coccobacillary organisms. These isolates were further confirmed to be *P. multocida* by various prescribed biochemical tests and were found to be positive for oxidase, catalase, indole production and nitrate reduction. The isolates were designated as P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, for P52 (vaccine strain), poultry, goat, buffalo and cattle isolates of *P. multocida*, respectively. Such an approach for isolation and characterization of *P. multocida* has also been used by Carter and Chengappa (1980) and Wijewardana et al. (1986).
4.1.1 Nucleic acid isolation

From pure culture of the organism, DNA was isolated according to method described by Wilson (1987) from all the isolates of *P. multocida*. The cultures were grown in BHI (Brain heart infusion) broth overnight at 37°C. Bacterial cells were harvested by centrifugation of broth culture and the pellet thus obtained was subjected to enzymatic digestion by proteinase-K-SDS method. Nucleic acid was extracted by standard phenol-chloroform method and precipitated using ethanol.

The qualitative and quantitative evaluation of DNA was carried out using agarose gel electrophoresis and NanoDrop Spectrophotometer ND 1000 (Thermo Scientific, USA). The Spectrophotometric evaluation of the extracted DNA revealed 680, 878, 843, 784 and 528 ng/μL concentration for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle respectively, as well as 260/280 ratio of 1.87, 1.93, 1.83, 1.89 and 1.92. Further, agarose gel electrophoresis was performed, which revealed fine quality genomic DNA (Plate 4.1).

4.1.2 PM-PCR based identification

For the PCR based identification and thereby confirmation of *P. multocida*, the protocol described by Townsend *et al.* (1998), specific for PM-PCR using KMT1SP6 and KMT1T7 primers were used with some minor modification. All the five isolates including P52 vaccine strain showed the expected amplified product of approximately 460bp, thus confirming them to be the isolates of *P. multocida* (Plate 4.2). Usefulness of PM-PCR using KMT1SP6 and KMT1T7 primers has also been reported earlier (Townsend *et al.*, 1998; Townsend *et al.*, 2001; Kro *et al.*, 2011; OIE, 2012).
Plate 4.1: Agarose gel electrophoresis of extracted genomic DNA of *P. multocida* isolates

Plate 4.2: PM PCR based identification of *P. multocida* isolates showing 460bp product size for all the isolates. M: Molecular marker (100bp), P: Positive control, N: Negative control
4.2 WHOLE GENOME SEQUENCING OF *P. multocida* ISOLATES

4.2.1 DNA nebulization, library quantity and quality assessment

4.2.1.1 Library quantity assessment using rapid library calculator

(Flurimeter assay)

DNA extracted from all the five isolates as in section 4.1.1 was used for the 454 GS FLX DNA library preparation. All the samples (DNA of the isolates) were nebulized and MIDs adaptor ligated as per the protocols in section 3.2.

Library quantitation was done by comparing sample with known concentration of the standard by preparing a standard curve (Plate 4.3), available at (http://454.com/my454/tools-downloads/rapid-library-calc.asp). Taking standard curve as the reference, reading of unknown sample was calculated as 7.27 X 10^8 molecules/μL, 5.85 X 10^8 molecules/μL, 7.93 X 10^8 molecules/μL, 3.93 X 10^8 molecules/μL and 3.01 X 10^8 molecules/μL for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle respectively.

Stock sample (dsDNA library) was diluted to get 1 X 10^7 molecules/μL by mixing stock library sample and TE Buffer (pH 8.0) as per Table 4.1.

**Table 4.1: Stock library preparation**

<table>
<thead>
<tr>
<th>Sample name</th>
<th>Stock library sample</th>
<th>TE buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>P52VAC</td>
<td>6.9 μL</td>
<td>493.1 μL</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>8.5 μL</td>
<td>491.5 μL</td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>6.3 μL</td>
<td>493.7 μL</td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>14.8 μL</td>
<td>485.2 μL</td>
</tr>
<tr>
<td>Anand1_cattle</td>
<td>16.6 μL</td>
<td>483.4 μL</td>
</tr>
</tbody>
</table>
4.2.1.2 Bioanalyser assay

For library quantitation, size selected dsDNA fragments were run on High Sensitivity DNA Chip in Bioanalyzer 2100 (Agilent Technologies, USA). Bioanalyzer results showed uniform distribution of DNA fragments across the length with average fragment size of 997bp, 892bp, 867bp, 783bp and 763bp as well as concentration of 2,706.62, 605.61, 822.70, 1,395.41 and 694.12 pg/μL for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively (Plate 4.4).

The DNA library, diluted to final concentration of $1 \times 10^6$ molecules/μL by taking 10μL of dsDNA and 90μL TE buffer (pH 8.0) for all the samples, denatured and finally used for emPCR. After small volume emPCR, beads were recovered, sequencing primers were attached and then percentage enrichment was calculated using following formula;

$$\% \text{ enriched beads} = \frac{\text{Number of final enriched beads}}{\text{Total number of beads}} \times 100$$

Tubes with obtained different molecules/beads ratio with % beads recovery details were as per Table 4.2.

**Table 4.2: Percentage bead recovery of small volume emPCR**

<table>
<thead>
<tr>
<th>Tube no.</th>
<th>Molecules/beads</th>
<th>No. of beads recovered</th>
<th>% bead recovery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tube 1</td>
<td>2.0</td>
<td>$0.72 \times 10^5$</td>
<td>3</td>
</tr>
<tr>
<td>Tube 2</td>
<td>4.0</td>
<td>$2.16 \times 10^5$</td>
<td>9</td>
</tr>
<tr>
<td>Tube 3</td>
<td>8.0</td>
<td>$3.84 \times 10^5$</td>
<td>16</td>
</tr>
<tr>
<td>Tube 4</td>
<td>16.0</td>
<td>$6.00 \times 10^5$</td>
<td>25</td>
</tr>
</tbody>
</table>
Plate 4.3: GS FLX library calculation for all the five isolates of *P. multocida*

Plate 4.4: Flurogram of a nebulized DNA sample after dsDNA library preparation
From percentage bead recovery of small volume (SV) emPCR, ideal bead recovery (i.e. 9.0%) having 4.0 molecules/bead was considered for LV (large volume) emulsion PCR. The clonally amplified DNA beads were enriched, annealed sequencing primers and finally DNA beads were used for sequencing run. On completion of sequencing run, raw data in form of images was transferred to GS FLX cluster via file transfer protocol (ftp), where signal (data) processing was done.

4.3 SEQUENCE ANALYSIS

After signal processing, reads were generated for all five *P. multocida* isolates and sequencing depth was calculated using below mentioned formula.

\[
\text{Sequencing depth (Coverage)} = \frac{\text{Total number of bases in reads}}{\text{Total number of bases in reference genome}}
\]

Whole genome sequence available for *P. multocida* Pm70 strain was used for mapping of all the five isolates using GS Reference Mapper version 2.3 and remaining reads which were not mapped, were assembled using GS De novo Assembler (Newbler) version 2.3 to check whether any read contains sequences for plasmid as well as for unique sequences related to *P. multocida*. Detailed results of mapping and assembly are given in Tables 4.3 and 4.4 respectively.

To confirm the results of mapping and assembly, clustal multiple sequence alignment of the isolates was also done with Pm70 reference strain sequence using MUSCLE. A representative clustal multiple alignment (for *HemU* DNA sequence) is shown in Plate 4.5, which shows sequence match for all the five isolates of *P. multocida* with Pm70. Currently available GS FLX Titanium instrument is having a capacity to sequence around 400 - 500bp at a time in 200 cycles of nucleotides (A, T, G and C) flow. As shown in Plate 4.6, most of the reads were in the size range of 300 - 500bp.
Plate 4.5: Clustal multiple alignment of all *P. multocida* isolates with Pm70 reference strain for *HemU* DNA sequence using MUSCLE

Plate 4.6: Read length distribution of a representative sample (P52VAC) of *P. multocida*
Table 4.3: Summary of data generated by GS FLX cluster after reference mapping with Pm70

<table>
<thead>
<tr>
<th>Mapping</th>
<th>P52VAC</th>
<th>Anand1_poultry</th>
<th>Anand1_goat</th>
<th>Anand1_buffalo</th>
<th>Anand1_cattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sequence reads</td>
<td>118843</td>
<td>113997</td>
<td>105729</td>
<td>134886</td>
<td>31346</td>
</tr>
<tr>
<td>Total Sequence Bases</td>
<td>42,598,100</td>
<td>29,000,497</td>
<td>21,890,353</td>
<td>39,756,349</td>
<td>7,429,658</td>
</tr>
<tr>
<td>Average Read Length</td>
<td>358</td>
<td>202</td>
<td>207</td>
<td>294</td>
<td>237</td>
</tr>
<tr>
<td>Coverage</td>
<td>18.87</td>
<td>12.85</td>
<td>9.70</td>
<td>17.61</td>
<td>3.29</td>
</tr>
<tr>
<td>Total Mapped Reads</td>
<td>105327</td>
<td>97674</td>
<td>95092</td>
<td>86765</td>
<td>24967</td>
</tr>
<tr>
<td>Total Mapped Reads (%)</td>
<td>88.63%</td>
<td>85.70%</td>
<td>89.95%</td>
<td>64.34%</td>
<td>79.67%</td>
</tr>
<tr>
<td>Total Mapped Bases</td>
<td>38,079,806</td>
<td>20,085,356</td>
<td>19,867,143</td>
<td>25,095,466</td>
<td>6,1451,56</td>
</tr>
<tr>
<td>Total Mapped Bases (%)</td>
<td>89.52%</td>
<td>87.38%</td>
<td>90.81%</td>
<td>63.22%</td>
<td>82.87%</td>
</tr>
<tr>
<td>Unmapped Reads</td>
<td>11308</td>
<td>12439</td>
<td>8743</td>
<td>42316</td>
<td>4820</td>
</tr>
<tr>
<td>Unmapped Reads (%)</td>
<td>9.52%</td>
<td>10.91%</td>
<td>8.270%</td>
<td>31.38%</td>
<td>15.38%</td>
</tr>
<tr>
<td>Total Contigs</td>
<td>155</td>
<td>402</td>
<td>257</td>
<td>227</td>
<td>1998</td>
</tr>
<tr>
<td>Total Contig Bases</td>
<td>2,069,867</td>
<td>2,024,810</td>
<td>2,097,193</td>
<td>2,064,639</td>
<td>1,819,413</td>
</tr>
<tr>
<td>Average. Contig Size</td>
<td>18716</td>
<td>10096</td>
<td>24768</td>
<td>1057</td>
<td>1339</td>
</tr>
<tr>
<td>Largest Contig</td>
<td>169703</td>
<td>88697</td>
<td>55720</td>
<td>90166</td>
<td>10792</td>
</tr>
<tr>
<td>N50* Contigs</td>
<td>44873</td>
<td>20025</td>
<td>23885</td>
<td>25302</td>
<td>1550</td>
</tr>
</tbody>
</table>

N50* is the length for which the collection of all contigs of that length or longer contains at least half of the sum of the lengths of all contigs.
Results and Discussion...

Table 4.4: Summary of data generated by GS FLX cluster after *De novo* assembly

<table>
<thead>
<tr>
<th>De Novo Assembly</th>
<th>P52VAC</th>
<th>Anand1_poultry</th>
<th>Anand1_goat</th>
<th>Anand1_buffalo</th>
<th>Anand1_cattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total unmapped reads</td>
<td>11308</td>
<td>12439</td>
<td>8743</td>
<td>42316</td>
<td>4820</td>
</tr>
<tr>
<td>Total unmapped bases</td>
<td>3,911,529</td>
<td>2,467,578</td>
<td>1,752,951</td>
<td>13,962,856</td>
<td>1,104,849</td>
</tr>
<tr>
<td>Total aligned reads</td>
<td>10775</td>
<td>10797</td>
<td>7287</td>
<td>24709</td>
<td>3002</td>
</tr>
<tr>
<td>Total aligned reads (%)</td>
<td>95.29%</td>
<td>62.28%</td>
<td>89.95%</td>
<td>58.40%</td>
<td>79.67%</td>
</tr>
<tr>
<td>Total aligned bases</td>
<td>3,834,503</td>
<td>2,197,589</td>
<td>1,490,730</td>
<td>7,805,067</td>
<td>7,522,44</td>
</tr>
<tr>
<td>Total aligned bases (%)</td>
<td>98.03%</td>
<td>68.10%</td>
<td>90.81%</td>
<td>55.90%</td>
<td>82.87%</td>
</tr>
<tr>
<td>Total Contigs</td>
<td>78</td>
<td>157</td>
<td>123</td>
<td>1359</td>
<td>192</td>
</tr>
<tr>
<td>Total contigs bases</td>
<td>207,091</td>
<td>213,470</td>
<td>192,527</td>
<td>1,235,377</td>
<td>143,000</td>
</tr>
<tr>
<td>Average contig size</td>
<td>2655</td>
<td>1360</td>
<td>1565</td>
<td>909</td>
<td>745</td>
</tr>
<tr>
<td>Largest contig size</td>
<td>34017</td>
<td>10692</td>
<td>17034</td>
<td>1057</td>
<td>3616</td>
</tr>
<tr>
<td>N50 contig</td>
<td>6306</td>
<td>2142</td>
<td>4255</td>
<td>969</td>
<td>1044</td>
</tr>
</tbody>
</table>
Results and Discussion...

In our experiment, all sequencing runs were filtered for individual samples based on MIDs tags attached at the time of adaptor ligation step, which resulted into 118,843, 113,997, 105,729, 134,886 and 31,346 reads with 42,598,100 (42.59Mb), 29,000,497 (29.00Mb), 21,890,353 (21.89Mb), 39,756,349 (39.75Mb) and 7,429,658 (7.42Mb) sequence bases for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. The coverage obtained were 18.87, 12.85, 9.70, 17.61 and 3.29 for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. All the isolates were mapped with Pm70 strain isolated from poultry as a reference strain (May et al., 2001) and generated 38,079,806 (89.52%), 20,085,356 (87.38%), 19,867,143 (90.81%), 25,095,466 (63.22%) and 6,145,156 (82.87%) mapped bases with 105,327, 97,674, 95,092, 86,765 and 24,967 mapped reads for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. Mapping results generated 155, 402, 257, 227 and 1998 contigs for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively (Table 4.3).

Remaining reads which were not mapped to the reference strain Pm70, were assembled using GS De novo Assembler (Newbler) version 2.3 which generated 10,775 (95.29%), 10,797 (62.28%), 7287 (89.95%), 24,709 (58.40%) and 3002 (79.67%) total aligned reads with 3,834,503 (98.03%), 2,197,589 (68.10%), 1,490,730 (90.81%), 7,805,067 (55.90%) and 752,244 (82.87%) total aligned bases for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively (Table 4.4). De novo assembly of reads were carried out for searching for plasmid. None of the De novo assembled reads matched to plasmid using BLAST hits.
Whole genome sequencing experiments were also carried out by May et al. (2001), Liu et al. (2012), Abrahante et al. (2013), Hauglund et al. (2013), Lainson et al. (2013\(^a\), Lainson et al. (2013\(^b\)), Yap et al. (2013), Abrahante et al. (2014); Davenport et al. (2014) and Vaid et al. (2014) for *P. multocida*. Since first whole genome sequence of *P. multocida* was carried out in 2001 when no whole genome sequencing platforms were available at that time, May et al. (2001) used Sanger sequencing approach. Afterward, as various whole genome sequencing platforms became available, the attempts for sequencing the whole bacterial genome increased. For Pm70, a poultry isolate, genome size is 2,257,487bp (2.25Mb) in length which is in a single scaffold as reported by May et al. (2001). After a decade of Pm70 sequencing, Liu et al. (2012) reported sequencing of HN06, a toxigenic *P. multocida* isolate of pig using Illumina Solexa platform and generated a total of 4,792,617 paired-end reads with ~360-fold coverage of the genome. *P. multocida* HN06 consists of a circular chromosome with 2,402,218bp (2.40Mb) and a plasmid, pHN06, with 5,360 bp.

Two virulent strains of *P. multocida*, P1059 and X73 of poultry were sequenced using 454 pyrosequencing by Abrahante et al. (2013). The assembly generated 2.30Mb and 2.26 Mbp genomes for P1059 and X73, respectively. *P. multocida* strain 671/90 belonging to A:3 serotype and isolated from infected calf lung tissue was sequenced by Lainson et al. (2103\(^b\)) using Illumina and 454 platforms, which produced 26 contigs with a genome size of 2,246,063bp (2.24Mb).

Whole genome sequencing for three bovine isolates (2000, RIIF, 1500E) from the United Kingdom and one bovine isolate from United States (P1933) using an Illumina Solexa Genome Analyzer was reported by Lainson et al. (2013\(^a\)). *De novo* assembly
Results and Discussion...

generated 261, 243, 265 and 269 contigs with 2,210,534bp (2.21Mb), 2,195,634bp (2.19Mb), 2,210,576bp (2.21Mb) and 2,441,686bp (2.44Mb) genome for 2000, RIIF, 1500E and P1933 isolates respectively.

PMTB strain isolated from buffalo carcass was sequenced using an Illumina genome analyzer and a total of 7,760,284 single-end reads were generated with approximately 100-fold genome coverage. PMTB has a genome size of 2,203,419bp (2.20Mb) as reported by Yap et al. (2013).

*P. multocida* subspecies *multocida* ATCC 43137, a type-A LPS strain commonly used as a reference strain in pathogenicity was studied by Davenport et al. (2014) using Illumina technology, which recorded 2.27Mb of genome size.

Whole genome sequencing of two strains, 2213 and 3213 of *P. multocida* isolated from the blood of carcasses of two buffaloes (*Bubalus bubalis*) by Abrahante et al. (2014) were carried out using the Illumina MiSeq platform. A total of 6,253,700 and 5,550,478 paired-end reads resulted into 34 and 32 contigs, respectively, for strains 2213 and 3213 with genome size of 2,309,333bp (2.30Mb) and 2,307,438bp (2.30Mb).

Another strain VTCCBAA264, a B:2 serotype of *P. multocida*, isolated from buffalo intestine was sequenced using 454 pyrosequencing by Vaid et al. (2014). A total of 123415 reads with ~23X coverage and 78 contigs with a total genome size of 2,280,332bp (2.28Mb) was obtained.

4.3.1 RAST analysis

For RAST analysis, all the reads of *P. multocida* were submitted to the RAST server available at (http://rast.nmpdr.org/rast.cgi). RAST produces two classes of asserted gene functions: subsystem-based assertions are based on recognition of functional
variants of subsystems, while non subsystem-based assertions are filled in using more common approaches based on integration of evidence from a number of tools. RAST distinguishes these two classes of annotation and uses the relatively reliable subsystem-based assertions as the basis for a detailed metabolic reconstruction (Aziz et al., 2008). Out of 27 RAST subsystems, as expected, subsystems for “Photosynthesis” and “Motility and Chemotaxis” were not assigned to any of the five isolates. The remaining 25 subsystems assigned with common genes/proteins for four isolates i.e. P52VAC, Anand1_poultry, Anand1_goat and Anand1_buffalo are shown in Table 4.5. Cattle isolate is omitted from this comparison as only 10 out of 27 subsystems were assigned, that too with lower counts because of the very less coverage obtained for this isolate during sequencing.

4.3.1.1 P52VAC

RAST analysis of P52VAC revealed genome size of 2,273,366bp with 209 contigs. Total number of subsystems defined was 419, while number of coding sequences were 2165 and number of RNAs were 68.

Based on subsystem analysis, out of 2165 coding sequences, 67% (1446 CDS) of genome was having defined subsystem based classification, while rest of the 33% (719 CDS) of genome was not having defined subsystem. Of the defined subsystem, highest abundance of subsystem were amino acids and derivatives (248 CDS), carbohydrates (244 CDS), protein metabolism (238 CDS), cofactor and vitamins, prosthetic groups, pigments (217 CDS) (Plate 4.7).

The most characteristic feature of P52VAC observed was that it contains colicin and bacteriocin production cluster which includes DedA, DedD and toxin. Bacteriocins
are antimicrobial peptides with different sizes, microbial targets and mechanisms of action produced by a large variety of bacteria. They are characterized by a bactericidal or bacteriostatic activity against strains of the same species or closely related species and differ from most therapeutic antibiotics due to their narrow activity spectrum and their proteinaceous nature (Zihler et al., 2009).

There was also presence of a gene cluster with resistance to fluoroquinolones which includes \textit{gyrA} and \textit{gyrB} genes coding for DNA gyrase as well as \textit{ParC} and \textit{ParD} genes coding for topoisomerase IV subunit. This implies to the antibiotic resistance shown by the bacteria against the most commonly employed antibiotics for the treatment of diseases caused by \textit{P. multocida}. Other important feature observed in the P52VAC strain is the presence of negative regulator of betalactamase expression, which is coded by \textit{BLR} gene cluster leading to resistance expressed by the organism. There was also presence of multidrug resistance efflux pump cluster of genes which were mainly \textit{MATE} (multidrug and toxin extrusion), \textit{MacA} and \textit{MacB} (macrolide specific efflux protein). Beside all these, genome of \textit{P. multocida} P52VAC contains all the proteins or genes required for the normal function for their survival. The common genes or proteins which have been shown to be of significance in \textit{P. multocida} or other bacterial spp. as per the available literature and found in our study as per RAST are listed in Table 4.5 with their potential role.
### Table 4.5 Coding sequences (CDS) in different functional groups in four isolates of *P. multocida*

<table>
<thead>
<tr>
<th>Functional group</th>
<th>Isolate*</th>
<th>No. of CDS</th>
<th>Genes/ Proteins</th>
<th>Function</th>
</tr>
</thead>
</table>

*Results and Discussion...*
<table>
<thead>
<tr>
<th>Virulence, Disease and Defense</th>
<th>P52VAC</th>
<th>47</th>
<th>DedA(+), toxin(+), PurF(+), gyrA, gyrB, parC, parE, BLR(+), TolC(+), MATE(+), MacA(+), MacB(+), MfrF(+), AcrB(+)</th>
<th>Colicin and bacteriocin production cluster, Resistance to fluoroquinolones, Betalactamase negative regulator, Multidrug efflux pump</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>54</td>
<td>DedA(+), toxin(+), PurF(+), gyrA, gyrB, parC, parE, BLR(+), TolC(+), MATE(+), MacA(+), MacB(+), MfrF(+), AcrB(+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>50</td>
<td>DedA(+), toxin(+), PurF(+), gyrA, gyrB, parC, parE, BLR(+), TolC(-), MATE(-), MacA(-), MacB(-), MfrF(-), AcrB(-)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>54</td>
<td>DedA(-), toxin(-), PurF(-), gyrA, gyrB, parC, parE, BLR(-), TolC(+), MATE(+), MacA(+), MacB(-), MfrF(-), AcrB(-)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Potassium metabolism</th>
<th>P52VAC</th>
<th>12</th>
<th>MSC(+), TkA, FKB, AATP, TrkG, KtrA, KtrB</th>
<th>Potassium homeostasis, Hyperosmotic potassium uptake</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>17</td>
<td>MSC(+), TkA, FKB, AATP, TrkG, KtrA, KtrB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>11</td>
<td>MSC(+), TkA, FKB, AATP, TrkG, KtrA, KtrB</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>14</td>
<td>MSC(-), TkA, AATP, FKB, TrkG, KtrA, KtrB</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Miscellaneous</th>
<th>P52VAC</th>
<th>35</th>
<th>Mei, Yje, QUEA, Tgt, HfIC(+), HfIK, DedA</th>
<th>Scaffold protein for [4Fe-4S] cluster assembly, DedA family inner membrane protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>35</td>
<td>Mei, Yje, QUEA, Tgt, HfIC(+), HfIK, DedA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>33</td>
<td>Mei, Yje, QUEA, Tgt, HfIC(+), HfIK, DedA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>37</td>
<td>Mei, Yje, QUEA, Tgt, HfIC(-), HfIK, DedA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phage, Prophage, Transposable elements, Plasmid</th>
<th>P52VAC</th>
<th>17</th>
<th>SpellVar(+), Rep(-), dPol 3(+), Hel (+), TrmlL(+), Port(+)</th>
<th>Phage packaging machinery, Phage replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>29</td>
<td>SpellVar(-), Rep(-), dPol 3(-), Hel (-), TrmlL(-), Port(-)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>51</td>
<td>SpellVar(+), Rep(+), dPol 3(+), Hel (+), TrmlL(+), Port(+)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>07</td>
<td>SpellVar(-), Rep(-), dPol 3(-), Hel (-), TrmlL(-), Port(-)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Membrane Transport</th>
<th>P52VAC</th>
<th>86</th>
<th>SapA(+), SapB(+), SapC(+), SapD(+), SapF(+), oppA, oppB, oppC, oppD(-), oppF, TonB(+), TonR, ExbB, ExdB, TolA, TolB, HTAS, pal, HasA</th>
<th>Peptide ABC transporter system Sap, ABC transported oligopeptide, Ton and Tol transport systems</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>86</td>
<td>SapA(+), SapB(+), SapC(+), SapD(+), SapF(+), oppA, oppB, oppC, oppD(-), oppF, TonB(+), TonR, ExbB, ExdB, TolA, TolB, HTAS, pal, HasA</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
-a
c
_c
cc

*cn

i_"^-c i-s-o

3

D.

oo-r- Q - e

c

-^,t

^ " S i^'S 0:2

lu

§-§6(5 2 S^

^

3

3^S

Si
Si.

a:

Q- b-i

2

ft!

ft;

1

1
OH

a;

,^ - 1^^ Q-^

££1

1

^ U

2

^_

^_- t t , B-H

_>

«5~

1"

5
ft!

ft!

ft!

ft!

^^

1 1 1 1

- /^

^ i2 ^-

I' , «

en ^

CO feT "^

J^ ft?

ftj

ft!

OH

as"

ft!"

1 1

^^
ft(

ft!

^^ 2
t^

as i*J 33

a< t<i a:;

QCi:3 kj

•R U "N

§

•S U <N

-a -a

bj §

^ft!

S.23

_tc
^ N
'^ ^C Sir

^ ft5 S

c^

_r N ^

tC

S.tc: (1,

ON

<

*•

00

ON

o

1< «*-

S

00

•*-»

o

O

Q.

c
c

<

-a

c

-

ON

_o

o

«

*^

S'ft;''
^ ^
S -

__l

<
>

T3
C
C8
C

in

<

<
B

E

O

w)

1 1
b
HH

u
R

B
C«

e
C

3
0

o
bo
_l
-a
c
CO
c

c
c

<

U
<
>
^/^

<

0-

C
CO
C

<
B
in

"o
X)

2S

CO
0

St
3

•a

c
CO
c

_l
-a
c
CO
c

<

<


<table>
<thead>
<tr>
<th><strong>Nucleosides and Nucleotides</strong></th>
<th><strong>Protein Metabolism</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P52VAC</strong></td>
<td><strong>Anand1_poultry</strong></td>
</tr>
<tr>
<td><strong>Anand1_poultry</strong></td>
<td><strong>Anand1_goat</strong></td>
</tr>
<tr>
<td><strong>Anand1_buffalo</strong></td>
<td><strong>Anand1_buffalo</strong></td>
</tr>
</tbody>
</table>

**Nucleosides and Nucleotides**


**Protein Metabolism**

<table>
<thead>
<tr>
<th>Cell Division and Cell Cycle</th>
<th>P52VAC</th>
<th>32</th>
<th>MreB(+), MreC(+), MreD(+), RodA(+), Maf(-), FisA(+), FisQ(+), ZipA(+), ZapA(+), MinC(+), IspA, YciA, SpoT, RelA</th>
<th>Bacterial cytoskeleton, Intracellular septation in enterobacteria, Stringent response ppGpp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>39</td>
<td></td>
<td>MreB(+), MreC(+), MreD(+), RodA(+), Maf(+), FisA(+), MraZ(+), FisI(+), FisQ(+), ZipA(+), ZapA(+), MinC(+), IspA, YciA, SpoT, RelA</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>33</td>
<td></td>
<td>MreB(+), MreC(+), MreD(+), RodA(+), Maf(+), FisA(+), MraZ(+), FisI(+), FisQ(+), ZipA(+), ZapA(+), MinC(+), IspA, YciA, SpoT, RelA</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>27</td>
<td></td>
<td>MreB(-), MreC(-), MreD(-), RodA(-), Maf(-), FisA(-), MraZ(-), FisI(-), FisQ(-), ZipA(-), ZapA(-), MinC(-), IspA, YciA, SpoT, RelA</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_buffalo</td>
<td>54</td>
<td></td>
<td>FNR, CGA(+), AC, PDE, APRT, Clp, LsrK(+), LsrR,</td>
<td></td>
</tr>
</tbody>
</table>

---

119
<table>
<thead>
<tr>
<th>Secondary Metabolism</th>
<th>Secondary Metabolism</th>
<th>LsrA, LsrC (-), LsrB, LsrD, LuxS, YcJx, UcJF</th>
<th>Results and Discussion...</th>
</tr>
</thead>
<tbody>
<tr>
<td>P52VAC</td>
<td>4</td>
<td>APRT, PRAI, Tsa, Tsb</td>
<td>Auxin biosynthesis</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>4</td>
<td>APRT, PRAI, Tsa, Tsb</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>5</td>
<td>APRT, PRAI, Tsa, Tsb</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA Metabolism</td>
<td>DNA Metabolism</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P52VAC</td>
<td>95</td>
<td>Cas1(+), Cas2(+), Csy2(+), Csy4(+), Csa1(+), Csa2(+), Csa3(+), Csa4(+), hyp1, dnaX, hyp2(+), recR, Topo-III</td>
<td>CRISPRs, DNA processing cluster</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>107</td>
<td>Cas1(-), Cas2(-), Csy2(-), Csy4(-), Csa1(-), Csa2(-), Csa3(-). hyp1, dnaX, hyp2(+), recR, Topo-III</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>101</td>
<td>Cas1(+), Cas2(+), Csy2(+), Csy4(+), Csa1(+), Csa2(+), Csa3(+), Csa4(+). hyp1, dnaX, hyp2(+), recR, Topo-III</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>134</td>
<td>Cas1(+), Cas2(+), Csy2(+), Csy4(+), Csa1(+), Csa2(+), Csa3(+), Csa4(+), hyp1, dnaX, hyp2(+), recR, Topo-III</td>
<td></td>
</tr>
<tr>
<td>Fatty Acids, Lipids and Isoprenoids</td>
<td>Fatty Acids, Lipids and Isoprenoids</td>
<td>3EE, 3EZ, DXS, ISPC, ISPd, ISPf, GTT, Upps, DMAT, Cdl, BCCP, BC, AccA, AccD, ACP, ACPS, FAB_FH(+). FabD, FabG, FabZ, FabB, FabA</td>
<td>Isoprenoids, Isoprenoid biosynthesis, Cardiopilin metabolism, Fatty acid biosynthesis FASII</td>
</tr>
<tr>
<td>P52VAC</td>
<td>78</td>
<td>3EE, 3EZ, DXS, ISPC, ISPd, ISPf, GTT, Upps, DMAT, Cdl, BCCP, BC, AccA, AccD, ACP, ACPS, FAB_FH(+). FabD, FabG, FabZ, FabB, FabA</td>
<td></td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>77</td>
<td>3EE, 3EZ, DXS, ISPC, ISPd, ISPf, GTT, Upps, DMAT, Cdl, BCCP, BC, AccA, AccD, ACP, ACPS, FAB_FH(+). FabD, FabG, FabZ, FabB, FabA</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>81</td>
<td>3EE, 3EZ, DXS, ISPC, ISPd, ISPf, GTT, Upps, DMAT, Cdl, BCCP, BC, AccA, AccD, ACP, ACPS, FAB_FH(+). FabD, FabG, FabZ, FabB, FabA</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>101</td>
<td>3EE, 3EZ, DXS, ISPC, ISPd, ISPf, GTT, Upps, DMAT, Cdl, BCCP, BC, AccA, AccD, ACP, ACPS, FAB_FH(+). FabD, FabG, FabZ, FabB, FabA</td>
<td></td>
</tr>
<tr>
<td>Nitrogen Metabolism</td>
<td>P52VAC</td>
<td>23</td>
<td>NiRes1, NiRes2, NaRper, Reg, GS, GlnE, GlnD, PII</td>
</tr>
<tr>
<td>-------------------</td>
<td>-------</td>
<td>---</td>
<td>----------------------------------</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>26</td>
<td>NiRes1, NiRes2, NaRper, Reg, GS, GlnE, GlnD, PII</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>26</td>
<td>NiRes1, NiRes2, NaRper, Reg, GS, GlnE, GlnD, PII</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>14</td>
<td>NiRes1, NiRes2, NaRper, Reg, GS, GlnE, GlnD, PII</td>
<td></td>
</tr>
<tr>
<td>Dormancy and Sporulation</td>
<td>P52VAC</td>
<td>2</td>
<td>SulA, SpoVC</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>2</td>
<td>SulA, SpoVC</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>2</td>
<td>SulA, SpoVC</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>3</td>
<td>SulA, SpoVC</td>
<td></td>
</tr>
<tr>
<td>Respiration</td>
<td>P52VAC</td>
<td>74</td>
<td>TrrA, TrrB, TrrC, TrrS(+), TrrR(+)</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>82</td>
<td>TrrA, TrrB, TrrC, TrrS(-), TrrR(-)</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>80</td>
<td>TrrA, TrrB, TrrC, TrrS(+), TrrR(+)</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>91</td>
<td>TrrA, TrrB, TrrC, TrrS(+), TrrR(+)</td>
<td></td>
</tr>
<tr>
<td>Stress Response</td>
<td>P52VAC</td>
<td>77</td>
<td>Fe-stress, Fr, Redox, H2O2(+), Fnr(+), AhpC, SOD, CspA(+), CspD(+), DnaJ(+), DnaK(+), GrpE(+), SnpB(+), Hsp15(+), RsmL(+), UrpA, UrpE</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>68</td>
<td>Fe-stress, Fr, Redox, H2O2(+), Fnr(+), AhpC, SOD, CspA(+), CspD(+), DnaJ(+), DnaK(+), GrpE(+), SnpB(-), Hsp15(-), RsmL(-), UrpA, UrpE</td>
<td></td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>86</td>
<td>Fe-stress, Fr, Redox, H2O2(+), Fnr(+), AhpC, SOD, CspA(+), CspD(+), DnaJ(+), DnaK(+), GrpE(+), SnpB(-), Hsp15(-), RsmL(-), UrpA, UrpE</td>
<td></td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>70</td>
<td>Fe-stress, Fr, Redox, H2O2(-), Fnr(-), AhpC, SOD, CspA(-), CspD(-), DnaJ(-), DnaK(-), GrpE(-), SnpB(-), Hsp15(-), RsmL(-), UrpA, UrpE</td>
<td></td>
</tr>
</tbody>
</table>
### Results and Discussion...

<table>
<thead>
<tr>
<th>Metabolism of Aromatic Compounds</th>
<th>P52VAC</th>
<th>8</th>
<th>HPAH(+), HPADO, HPAHb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>8</td>
<td></td>
<td>HPAH(+), HPADO, HPAHb</td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>8</td>
<td></td>
<td>HPAH(+), HPADO, HPAHb</td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>22</td>
<td></td>
<td>HPAH(-), HPADO, HPAHb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino Acids and Derivatives</th>
<th>P52VAC</th>
<th>248</th>
<th>CysE, CysK, CysZ, CysB, CysT, HSDH, HSAT, HS, CTGS, CTBL, MS, MTHFR, MT, SAMS, SAHCN, RHMCCysE, CysK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>267</td>
<td></td>
<td>CysE, CysK, CysZ, CysB, CysT(--), HSDH, HSAT, HS, CTGS, CTBL, MS, MTHFR, MT, SAMS, SAHCN, RHMCCysE, CysK</td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>254</td>
<td></td>
<td>CysE, CysK, CysZ, CysB, CysT, HSDH, HSAT, HS, CTGS, CTBL, MS, MTHFR, MT, SAMS, SAHCN, RHMCCysE, CysK</td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>337</td>
<td></td>
<td>CysE, CysK, CysZ, CysB, CysT, HSDH, HSAT, HS, CTGS, CTBL, MS, MTHFR, MT, SAMS, SAHCN, RHMCCysE, CysK</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sulfur Metabolism</th>
<th>P52VAC</th>
<th>19</th>
<th>DsrE, DsrH(+), DsrC-4, TcyP, MetI, MetC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand1_poultry</td>
<td>18</td>
<td></td>
<td>DsrE, DsrH(+), DsrC-4, TcyP, MetI, MetC</td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>17</td>
<td></td>
<td>DsrE, DsrH(+), DsrC-4, TcyP, MetI, MetC</td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>17</td>
<td></td>
<td>DsrE, DsrH(-), DsrC-4, TcyP, MetI, MetC</td>
</tr>
</tbody>
</table>

|----------------------------------|--------|-----|----------------------------------|

Aromatic amin catabolism

Cystein biosynthesis, Methionine biosynthesis

Sulfite reduction-associated complex, L-cystein uptake and metabolism

Phosphate metabolism
### Carbohydrates


| Carbohydrates | P52VAC | 244 | XylA(+), XylB, Xyl_T(+), Xyn_T(+), mdh(+), gck(+), eno(+), MTFC(+), GlyA(+), sdbh(+), sdaA(+), scaA(+), glutA(+), LldD, LidEFG, LidP | Xylose utilization, Serine-glyoxanate cycle, Lactate utilization |
| Anand1_poultry | 237 | XylA(-), XylB, Xyl_T(-), Xyn_T(-), mdh(-), gck(-), eno(-), MTFC(-), GlyA(-), sdbh(-), sdaA(-), scaA(-), glutA(-), LldD, LidEFG, LidP |  |
| Anand1_goat | 243 | XylA(-), XylB, Xyl_T(-), Xyn_T(-), mdh(+), gck(+), eno(+), MTFC(+), GlyA(+), sdbh(+), sdaA(+), scaA(+), glutA(+), LldD, LidEFG, LidP |  |
| Anand1_buffalo | 239 | XylA(-), XylB, Xyl_T(+), Xyn_T(+), mdh(-), gck(-), eno(-), MTFC(-), GlyA(-), sdbh(-), sdaA(-), scaA(-), glutA(-), LldD, LidEFG, LidP |  |

- Cattle isolate Anand1_cattle is omitted due to very less coverage obtained.
- Genes/proteins in bold letters indicate their common presence in all the four isolates.
- Genes/proteins with (+) sign indicate their presence in one or more isolates.
- Genes/proteins with (-) sign indicate their absence in one or more isolates.
4.3.1.2 Anandl_poultry

RAST analysis of Anandl_poultry revealed genome size of 2,227,943bp with 489 contigs. Total number of subsystem defined was 406, while number of coding sequences (CDS) were 2365 and number of RNAs were 54.

Based on subsystem analysis, 65% (1529 CDS) of genome was having defined subsystem based classification, while rest of the 35% (836 CDS) of genome was not having defined subsystem. Of the defined subsystem, highest abundance of subsystem were Amino acids and derivatives (267 CDS), carbohydrates (237 CDS), cofactor and vitamins, prosthetic groups, pigments (235 CDS), and protein metabolism (224 CDS) (Plate 4.8).

The genome of Anandl_poultry contains genes for resistance to antibiotics and toxic compounds, which includes lysozyme inhibitors, copper homeostasis, cobalt-zinc-cadmium resistance, arsenic resistance, copper homeostasis: copper tolerance, resistance to fluoroquinolones, beta-lactamase and multidrug resistance efflux pumps. The multidrug resistance efflux pumps confer resistance to natural substances produced by the host, including bile, hormones and host-defence molecules. In addition, some efflux pumps of the resistance nodulation division (RND) family have been shown to have a role in the colonization and the persistence of bacteria in the host. So, they have roles in bacterial pathogenicity. Genes for resistance to fluoroquinolones include ParC, ParE which code for topoisomerase IV subunit as well as gyrA and gyrB, which code for DNA gyrase subunit B and lead to the resistance of bacteria against commonly used fluoroquinolone group of antibiotics against Gram negative bacteria. The list of genes required for the normal survival, replication, protein synthesis as well as for bacterial pathogenicity are listed in Table 4.5 as per RAST server.
Plate 4.7: Subsystem distribution of P52VAC isolate of *P. multocida* (based on RAST annotation server)

Plate 4.8: Subsystem distribution of Anand1_poultry isolate of *P. multocida* (based on RAST annotation server)
4.3.1.3 Anand1_goat

RAST analysis of Anand1_goat revealed genome size of 2,285,382bp with 349 contigs. Total number of subsystems defined was 419, number of coding sequences (CDS) were 2340 and number of RNAs were 54.

Based on subsystem analysis, 66% (1543 CDS) of genome was having defined subsystem based classification, while rest of the 34% (797 CDS) of genome was not having defined subsystem. Of the defined subsystem, highest abundance of subsystem were Amino acids and derivatives (254 CDS), protein metabolism (252 CDS), carbohydrates (243 CDS), and cofactor and vitamins, prosthetic groups, pigments (242 CDS) (Plate 4.9).

For Anand1_goat isolate, important feature is the presence of periplasmic hemin binding protein (HBP and ExB, TonB). Hemin is the oxidized form of heme that is found in extracellular environments, and it is the form of heme transported by bacterial uptake systems. Hemin iron transport and utilization systems have been identified in numerous bacterial species and were initially described in gram-negative pathogens. It is an outer membrane receptor and a periplasmic binding protein-dependent ABC-type transporters required for hemin uptake. Certain Gram negative species also obtain hemin iron through the use of hemophores, which are low-molecular-weight secreted hemin binding proteins that are able to extract hemin from hemoglobin and then transfer the hemin to receptors in the bacterial outer membrane (Allen and Schmitt, 2009).

There was also presence of Hemin ABC transporter particularly permease protein, Arsenic efflux pump protein, Multidrug resistance transporter, Ber/CflA family, Multi antimicrobial extrusion protein (Na(+)/drug antiporter), MATE family of MDR efflux pumps, resistance-nodulation-cell division superfamily (RND) efflux
system, outer membrane lipoprotein CmeC in the genome of Anand1_goat isolate. The genetic elements encoding efflux pumps may be encoded on chromosomes and/or plasmids, thus contributing to both intrinsic (natural) and acquired resistance respectively. As an intrinsic mechanism of resistance, efflux pump genes can survive a hostile environment (e.g. in the presence of antibiotics) which allows for the selection of mutants that over-express these genes. Being located on transportable genetic elements as plasmids or transposons is also advantageous for the microorganisms as it allows for the easy spread of efflux genes between distant species. Antibiotics can act as inducers and regulators of the expression of some efflux pumps (Morita et al., 2006). Expression of several efflux pumps in a given bacterial species may lead to a broad spectrum of resistance when considering the shared substrates of some multi-drug efflux pumps, where one efflux pump may confer resistance to a wide range of antimicrobials. The major facilitator superfamily (MFS) mainly present in the Gram positive bacteria was not present in any of the isolates studied in present study. The list of genes found present in the genome of Anand1_goat are listed in Table 4.5 as per RAST server.

4.3.1.4 Anand1_buffalo

RAST analysis of Anand1_buffalo revealed genome size of 2,045,610bp with 2188 contigs. Total number of subsystems defined was 345, number of coding sequences (CDS) were 2680 and number of RNAs were 40.

Based on subsystem analysis, 62% (1656 CDS) of genome was having defined subsystem based classification, while rest of the 38% (1024 CDS) of genome was not having defined subsystem. Of the defined subsystem, highest abundance of subsystem were amino acids and derivatives (357 CDS), cofactor and vitamins, prosthetic
groups, pigments (294 CDS), carbohydrates (239 CDS), and protein metabolism (200 CDS) (Plate 4.10).

The main distinguishing feature of Anand1_buffalo isolate was the presence of
*Mycobacterium* virulence operon involved in protein synthesis. Additionally, there
was presence of “Heme, hemin uptake and utilization systems in Gram negatives”
which includes *ExB* gene, involved in ferric siderophore transport system, *FCR* which
codes for the TonB-dependent hemin, ferrichrome receptor *RP2*, which is an outer
membrane receptor protein mostly involved in Fe transport. There was also presence
of *HBP*, which is periplasmic heme binding protein and *HemR* gene coding for hemin
receptor protein. Apart from all these, genes involve in protein metabolism, synthesis
as well as for normal functions of bacteria were found as listed in Table 4.5.

4.3.1.5 Anand1_cattle

RAST analysis of Anand1_cattle revealed genome size of 1,438,517bp with
3152 contigs. Total number of subsystems defined was 19, number of coding
sequences (CDS) were 1862 and number of RNAs were 0.

Based on subsystem analysis, 2% (32 CDS) of genome was having defined
subsystem based classification, while rest of the 98% (1830 CDS) of genome was not
having defined subsystem. This was due to less coverage obtained during sequencing
of the isolate Anand1_cattle i.e. 3.29X. Out of 27 subsystems, only 10 were assigned
and that too with lower counts. Hence, the subsystems assigned are very less. Of the
defined subsystem, highest abundance of subsystems were cofactor and vitamins,
prosthetic groups, pigments (14 CDS), amino acids and derivatives (9 CDS) and fatty
acid, lipids and isoprenoids (9 CDS) (Plate 4.11). Possible reason of less coverage
particularly for Anand1_cattle might be increased number of “mixed beads” or
loading of probably less number of beads at the time of sequencing. Second
Subsystem Statistics

<table>
<thead>
<tr>
<th>Subsystem Coverage</th>
<th>Subsystem Category Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Coverage" /></td>
<td><img src="image" alt="Distribution" /></td>
</tr>
</tbody>
</table>

**Subsystem Feature Counts**

- **Cofactors, Vitamins, Prosthetic Groups, Pigments (242)**
- **Cell Wall and Capsule (152)**
- **Virulence, Disease and Defense (50)**
- **Potassium metabolism (11)**
- **Photosynthesis (0)**
- **Miscellaneous (33)**
- **Phages, Prophages, Transposable elements, Plasmids (51)**
- **Membrane Transport (96)**
- **Iron acquisition and metabolism (69)**
- **RNA Metabolism (181)**
- **Nucleosides and Nucleotides (84)**
- **Protein Metabolism (252)**
- **Cell Division and Cell Cycle (33)**
- **Motility and Chemotaxis (0)**
- **Regulation and Cell signaling (43)**
- **Secondary Metabolism (5)**
- **DNA Metabolism (101)**
- **Fatty Acids, Lipids, and Isoprenoids (81)**
- **Nitrogen Metabolism (26)**
- **Dormancy and Sporulation (2)**
- **Respiration (80)**
- **Stress Response (86)**
- **Metabolism of Aromatic Compounds (8)**
- **Amino Acids and Derivatives (254)**
- **Sulfur Metabolism (17)**
- **Phosphorus Metabolism (23)**
- **Carbohydrates (243)**

Plate 4.9: Subsystem distribution of Anand1_goat isolate of *P. multocida* (based on RAST annotation server)

Subsystem Statistics

<table>
<thead>
<tr>
<th>Subsystem Coverage</th>
<th>Subsystem Category Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Coverage" /></td>
<td><img src="image" alt="Distribution" /></td>
</tr>
</tbody>
</table>

**Subsystem Feature Counts**

- **Cofactors, Vitamins, Prosthetic Groups, Pigments (294)**
- **Cell Wall and Capsule (179)**
- **Virulence, Disease and Defense (54)**
- **Potassium metabolism (14)**
- **Photosynthesis (0)**
- **Miscellaneous (37)**
- **Phages, Prophages, Transposable elements, Plasmids (7)**
- **Membrane Transport (106)**
- **Iron acquisition and metabolism (25)**
- **RNA Metabolism (139)**
- **Nucleosides and Nucleotides (117)**
- **Protein Metabolism (200)**
- **Cell Division and Cell Cycle (27)**
- **Motility and Chemotaxis (0)**
- **Regulation and Cell signaling (54)**
- **Secondary Metabolism (0)**
- **DNA Metabolism (134)**
- **Fatty Acids, Lipids, and Isoprenoids (101)**
- **Nitrogen Metabolism (14)**
- **Dormancy and Sporulation (3)**
- **Respiration (91)**
- **Stress Response (70)**
- **Metabolism of Aromatic Compounds (22)**
- **Amino Acids and Derivatives (337)**
- **Sulfur Metabolism (17)**
- **Phosphorus Metabolism (32)**
- **Carbohydrates (239)**

Plate 4.10: Subsystem distribution of Anand1_buffalo isolate of *P. multocida* (based on RAST annotation server)
Plate 4.11: Subsystem distribution of Anand1_cattle isolate of *P. multocida* (based on RAST annotation server)

Plate 4.12: G+C content of the genome for all the five isolates of *P. multocida*
possibility may be related with improper adaptor ligation at the time of library preparation, which may ultimately leads to decreased amplification of DNA sequences during emPCR (454 GS Run Browser manual available at http://454.com/my454/tools-downloads/software-downloads.asp).

4.3.2 PGAAP analysis

For Prokaryotic Genome Automatic Annotation Pipeline (PGAAP) analysis, all the reads were submitted to the PGAAP annotation pipeline of NCBI (National Center for Biotechnology Information) after removing reads which were less than 200bp (using in-house perlscript programme). Genomic statistics of all *P. multocida* isolates obtained by PGAAP analysis are given in Table 4.6. All the sequences were submitted to Genebank NCBI for open access under Genebank project id ALBZ0000000.1, AFRR0000000.1, AFRS0000000.1, ALBX0000000.1 and ALBY0000000.1 for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively.

Comparative G+C content of the five isolates obtained by PGAAP analysis are shown in Table 4.6 and Plate 4.12. Number of CDS and PEG of all the isolates are depicted in Table 4.6 and Plate 4.13 based on PGAAP analysis. Representative graph (for Anand1_goat isolate) showing number of contigs number and contig length is shown after annotation using PGAAP pipeline in Plate 4.14.

Whole genome sequence comparison of the five isolates of *P. multocida* and Pm70 genome sequences using progressiveMauve alignment showing collinear blocks of genes is represented in Plate 4.15. Rectangles of similar colour shade show collinear blocks of genes. Areas of low identity within colinear blocks are shown by reduced height of shading.
Plate 4.13: Graph showing number and length of contigs for representative sample (Anand1_goa isolate of *P. multocida*).

Plate 4.14: Number of CDS and PEG for all the isolates of *P. multocida*. 

**No. of CDS and PEG**

- **CDS (Coding Sequences)**
- **PEG (Protein Coding Genes)**

<table>
<thead>
<tr>
<th>Isolate</th>
<th>CDS</th>
<th>PEG</th>
</tr>
</thead>
<tbody>
<tr>
<td>P52VAC</td>
<td>2064</td>
<td>2194</td>
</tr>
<tr>
<td>Anand1_poultry</td>
<td>2337</td>
<td>2284</td>
</tr>
<tr>
<td>Anand1_goat</td>
<td>2319</td>
<td>2260</td>
</tr>
<tr>
<td>Anand1_buffalo</td>
<td>3258</td>
<td>3218</td>
</tr>
<tr>
<td>Anand1_cattle</td>
<td>3623</td>
<td>3590</td>
</tr>
</tbody>
</table>
Plate 4.15: Comparison of the five isolates of *P. multocida* and Pm70 genome sequences using progressive Mauve alignment (Darling et al., 2010)

Plate 4.16: Phylogenetic comparison among *Pasteurella multocida* species whole genome (complete and incomplete) sequences using MEGA6
Table 4.6: Genomic statistics of *P. multocida* isolate of different species obtained by PGAAP analysis

<table>
<thead>
<tr>
<th></th>
<th>P52VAC</th>
<th>Anand1_poultry</th>
<th>Anand1_goat</th>
<th>Anand1_buffalo</th>
<th>Anand1_cattle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>2.27Mb</td>
<td>2.23Mb</td>
<td>2.29Mb</td>
<td>2.05Mb</td>
<td>1.44Mb</td>
</tr>
<tr>
<td>INSDC</td>
<td>ALBZ00000000.1</td>
<td>AFRR00000000.1</td>
<td>AFRS00000000.1</td>
<td>ALBX00000000.1</td>
<td>ALBY00000000.1</td>
</tr>
<tr>
<td>Coverage</td>
<td>18.87X</td>
<td>12.85X</td>
<td>9.70X</td>
<td>17.61X</td>
<td>3.29X</td>
</tr>
<tr>
<td>Contig/ Scaffold</td>
<td>209</td>
<td>489</td>
<td>349</td>
<td>2188</td>
<td>3152</td>
</tr>
<tr>
<td>Bioproject</td>
<td>PRJNA 180802</td>
<td>PRJNA 179832</td>
<td>PRJNA 179833</td>
<td>PRJNA 180800</td>
<td>PRJNA 180801</td>
</tr>
<tr>
<td>Taxonomic id</td>
<td>1077143</td>
<td>1032866</td>
<td>1032869</td>
<td>1032868</td>
<td>1032867</td>
</tr>
<tr>
<td>Capsule</td>
<td>B (B:2)</td>
<td>A</td>
<td>D</td>
<td>B (B:2)</td>
<td>B (B:2)</td>
</tr>
<tr>
<td>G+C content</td>
<td>40.40%</td>
<td>40.20%</td>
<td>40.50%</td>
<td>40.90%</td>
<td>41.00%</td>
</tr>
<tr>
<td>Coding region</td>
<td>2,066</td>
<td>2,337</td>
<td>2,319</td>
<td>3,258</td>
<td>3,623</td>
</tr>
<tr>
<td>(CDS)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA genes</td>
<td>64</td>
<td>53</td>
<td>53</td>
<td>41</td>
<td>33</td>
</tr>
<tr>
<td>Protein coding</td>
<td>2,194</td>
<td>2,284</td>
<td>2,266</td>
<td>3,218</td>
<td>3,590</td>
</tr>
<tr>
<td>genes (PEG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N50</td>
<td>36,108</td>
<td>17,120</td>
<td>21,172</td>
<td>1,299</td>
<td>528</td>
</tr>
<tr>
<td>L50*</td>
<td>18</td>
<td>36</td>
<td>32</td>
<td>474</td>
<td>865</td>
</tr>
</tbody>
</table>

*L50* is the number of scaffolds that accounts for more than 50% of the genome assembly.
Phylogenetic analysis is a useful measure for deciphering the genetic relationship among the different isolates of bacterial spp. For phylogenetic analysis, available whole genome sequences of *P. multocida* (complete and incomplete) in FASTA files were downloaded from (http://www.ncbi.nlm.nih.gov/genome/genomes/912?). A phylogenetic tree based on whole genome was constructed using CV tree (Zhao and Hao, 2009) and results were visualized in MEGA6 (Tamura et al., 2013). Phylogenetic analysis revealed least genetic distance between isolates P52VAC and Anandl_buffalo, which grouped under the same cluster as compared to others. Anandl_cattle was found to be closely related to the isolate P52VAC and Anandl_buffalo. Anandl_poultry and Anandl_goat grouped in same cluster, however at a genetic distance from rest of the three isolates. Pm70 (poultry isolate) and PMTB (buffalo isolate) are closely related to our five isolates as compared to other isolates/strains from various parts of the world (Plate 4.16). Circular genome maps for all the five isolates of *P. multocida* were created using CGView server (Grant and Stothard, 2008) and are shown in Plates 4.17 to 4.21).

### 4.3.2.1 Virulence genes of *P. multocida*

Productive infection by bacterial pathogens relies on the expression of virulence factors that have wide ranging functions like competence, adherence, capsule synthesis and export, evading host immune responses etc. For all the isolates of *P. multocida* (i.e. P52VAC, Anandl_poultry, Anandl_goat, Anandl_buffalo and Anandl_cattle), gene locus or protein id were downloaded from http://www.ncbi.nlm.nih.gov/genome/genomes/912? under ‘Protein’ column/section. After downloading protein or gene locus individually, each were searched for the virulence associated genes manually falling under seven broad categories of virulence associated genes viz. capsule, fimbriae and adhesion, iron metabolism, outer
Plate 4.17: Circular genome map of P52VAC isolate of *P. multocida*.*
The map represents the circular view of the genome sequence of P52VAC of *P. multocida* (Mapped Contigs). The circle was created using CGView server for map creation. The outermost circular ring shows coding sequences (CDS) followed by tRNA and rRNA respectively. Circle 4 shows the G+C content. The innermost ring towards the center represents the GC skew information in the (+) strand (green color) and (-) strand (dark pink color). *Reads were mapped to reference genome using Genebank file and consensus sequence was created. Finally annotated consensus sequence with reference sequence was used to create circular map of genome.*

Plate 4.18: Circular genome map of Anand1_poultry isolate of *P. multocida*.*
The map represents the circular view of the genome sequence of Anand1_poultry *P. multocida* (Mapped Contigs). The circle was created using CGView server for map creation. The outermost circular ring shows coding sequences (CDS) followed by tRNA and rRNA respectively. Circle 4 shows the G+C content. The innermost ring towards the center represents the GC skew information in the (+) strand (green color) and (-) strand (dark pink color). *Reads were mapped to reference genome using Genebank file and consensus sequence was created. Finally annotated consensus sequence with reference sequence was used to create circular map of genome.*
Plate 4.19: Circular genome map of Anandl_goat isolate of P. multocida*

The map represents the circular view of the genome sequence of Anandl_goat P. multocida (Mapped Contigs). The circle was created using CGView server for map creation. The outermost circular ring shows coding sequences (CDS) followed by tRNA and rRNA respectively. Circle 4 shows the G+C content. The innermost ring towards the center represents the GC skew information in the (+) strand (green color) and (-) strand (dark pink color). *Reads were mapped to reference genome using Genebank file and consensus sequence was created. Finally annotated consensus sequence with reference sequence was used to create circular map of genome.

Plate 4.20: Circular genome map of Anandl_buffalo isolate of P. multocida*

The map represents the circular view of the genome sequence of Anandl_buffalo P. multocida (Mapped Contigs). The circle was created using CGView server for map creation. The outermost circular ring shows coding sequences (CDS) followed by tRNA and rRNA respectively. Circle 4 shows the G+C content. The innermost ring towards the center represents the GC skew information in the (+) strand (green color) and (-) strand (dark pink color). *Reads were mapped to reference genome using Genebank file and consensus sequence was created. Finally annotated consensus sequence with reference sequence was used to create circular map of genome.
Plate 4.21: Circular genome map of Anand1_cattle isolate of *P. multocida*

The map represents the circular view of the genome sequence of Anand1_cattle *P. multocida* (Mapped Contigs). The circle was created using CGView server. The outermost circular ring shows coding sequences (CDS) followed by tRNA and rRNA respectively. Circle 4 shows the G+C content. The innermost ring towards the center represents the GC skew information in the (+) strand (green color) and (-) strand (dark purple color). *Reads were mapped to reference genome using Genebank file and consensus sequence was created. Finally annotated consensus sequence with reference sequence was used to create circular map of genome.*
membrane protein, superoxide dismutase, sialic acid metabolism and transcription regulation. Out of these seven categories, all the five genes falling under three categories i.e. SodA and SodC under superoxide dismutase, NanH and NanB under sialic acid metabolism and Fis under transcription regulation category were present in all the five isolates (Tables 4.7 and 4.8).

Presence or absence of virulence associated genes under rest of the four categories viz. capsule, fimbriae and adhesion, iron metabolism and outer membrane protein was as follow.

A) Genes for capsule

Out of nine genes under this category, PglA and Kmt1 were found in all the five isolates, while HyaE was found to be present only in goat isolate. HexA and HexC were absent in buffalo and cattle isolates, while HexB and HexD were absent in goat and cattle isolates. KpsF gene was found to be absent in poultry and cattle isolates. Gene LctP was present only in goat and cattle isolates. Out of nine genes falling under this category, P52VAC and goat isolates revealed maximum number of seven genes, while buffalo isolate revealed only three out of nine genes in this category.

B) Genes for fimbriae and adhesion

Out of 16 genes under this category, Hsf, Pfhl2, Pfhr, PfJb, PlpB, and Plp4 were found in all the five isolates of P. multocida studied, which indicated the importance of these genes in virulence. HofC gene was absent in P52VAC only, whereas PlpE was absent in cattle as well as buffalo isolate. ComE was absent in P52VAC isolate, while TadE was absent in buffalo isolate. Pfhl1 and PlpP genes were absent in cattle isolate, while RcpA and RcpB were absent in buffalo isolate.
Table 4.7: Virulence associated genes of *P. multocida* from different species of animals by PGAAP analysis

<table>
<thead>
<tr>
<th>Sr. No.</th>
<th>Gene locus / Protein</th>
<th>P52VAC</th>
<th>Anand1_poultry</th>
<th>Anand1_goat</th>
<th>Anand1_buffalo</th>
<th>Anand1_cattle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Capsule</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>HexA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>HexB</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>HexC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>HexD</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>PglA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>Kmt1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>KpsF</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>HyaE</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>LctP</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td><strong>Fimbriae and adhesion</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ComE</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>11</td>
<td>ClpB</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>HofC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>Hsf</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>14</td>
<td>PfbB1</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>PfbB2</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>PfhR</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>17</td>
<td>PfB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>18</td>
<td>PlpB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>19</td>
<td>PlpE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>PlpP</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>RcpA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>22</td>
<td>RcpB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>
Results and Discussion...

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>23</td>
<td>TadE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>TadF</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>Plp4</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Iron metabolism</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>ExbB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>FbpA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>FbpB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>29</td>
<td>FbpC</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>30</td>
<td>HbpA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>31</td>
<td>HgbA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>32</td>
<td>HemR</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>33</td>
<td>HemU</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>34</td>
<td>OmpW</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>35</td>
<td>Rfb</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>36</td>
<td>RffG</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>37</td>
<td>TbpA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>38</td>
<td>TonB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>39</td>
<td>TonB-dependent lactoferrin and transferrin receptor</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Outer membrane protein (OMP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>TonB-dependent receptor</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>41</td>
<td>Translocation protein TolB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>42</td>
<td>HasR</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>LppB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>44</td>
<td>LspB</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>45</td>
<td>OmpH</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>46</td>
<td>Outer membrane lipoprotein LolB</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>47</td>
<td>Filamentous hemagglutinin outer membrane protein (PfH4)</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>48</td>
<td>PtfA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>49</td>
<td>Oma87</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>50</td>
<td>VacJ</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

**Superoxide dismutase**

| 51 | SodA | + | + | + | + | + | + |
| 52 | SodC | + | + | + | + | + | + |

**Sialic acid metabolism**

| 53 | NanB | + | + | + | + | + | + |
| 54 | NanH | + | + | + | + | + | + |

**Transcription regulation**

| 55 | Fis | + | + | + | + | + | + |

(+) indicate presence of gene/protein in the particular isolate
(-) indicate absence of gene/protein in the particular isolate

**Table 4.8: Unique virulence associated genes (Categorywise) in different isolates of *P. multocida***

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Found only in isolate of</th>
<th>Absent only in isolate of</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolate name</td>
<td>P52</td>
<td>Poultry</td>
</tr>
<tr>
<td>Capsule</td>
<td>HexA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HexB</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HexC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>HexD</td>
<td></td>
</tr>
</tbody>
</table>

P52
<table>
<thead>
<tr>
<th></th>
<th>KpsF</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HyaE</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LctP</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fimbriae and adhesion**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ComE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ClpB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HofC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PfhB1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PlpE</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PlpP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RcpA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RcpB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TadE</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TadF</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Iron metabolism**

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FbpA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>FbpC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>HemR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ThpA</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>TonB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>TonB- dependent lactoferrin and Transferrin receptor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>TonB- dependent receptor</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>TolB</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Translocation protein TolB</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Outer membrane protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipoprotein LolB</td>
</tr>
<tr>
<td>PfhA</td>
</tr>
<tr>
<td>Oma87</td>
</tr>
<tr>
<td>VacJ</td>
</tr>
</tbody>
</table>
ClpB was found to be absent in P52VAC and cattle isolates, whereas TadF was found to be absent in buffalo and cattle isolates. Out of 16 genes falling under “fimbriae and adhesion” category, P52 and goat isolates were having all the 16 genes present in their genome, while cattle isolate was having only 10 genes present out of 16 genes.

C) Genes for iron metabolism

Out of 16 genes falling under “iron metabolism” category, ExbB, FbpB, HbpA, HgbA, HemU, OmpW, Rfb and Rfg were present in all the five isolates studied. Genes FbpA and TonB were absent in cattle isolate, ThpA and TonB-dependent lactoferrin and transferrin receptor were absent in goat isolate. Gene FbpC was absent in buffalo isolate. TonB-dependent receptor was found to be present in poultry and goat isolates only. Translocation protein TolB was found to be present and HemR was absent only in P52VAC strain, which is a unique finding of this isolate. Under this category, maximum genes were present in poultry isolate, having 15 out of 16 genes studied with absence of only translocation protein TolB. Least number of genes were 12 for cattle isolate, which might be due to less sequencing depth/coverage.

D) Genes for outer membrane protein

The category of “outer membrane proteins (OMPs)” included nine genes, out of which, HasR, LppB, LspB, OmpH and PtfA were found in all the five isolates of P. multocida. Outer membrane protein LolB was found in P52VAC as well as in cattle isolate. But, the sequence of LolB gene was found to be truncated for the cattle isolate. Hence, genes LolB and PfhA were unique to the P52VAC strain. VacJ gene, which has been recently identified as a potential virulence factor, was absent in goat and cattle isolates. Oma87 was absent only in poultry isolate. PfhA gene codes for filamentous haemagglutinin outer membrane protein, which was present only in
Results and Discussion...
P52VAC isolate. Out of nine genes falling in the category of “outer membrane protein”, P52VAC revealed all the genes present in their genome, while poultry and goat isolates were having least number of six genes present out of nine genes.

Considering the literature available particularly for the virulence associated genes in *P. multocida*, following genes are described and discussed for their possible role in virulence or their association in the disease process and progression.

4.3.2.1.1 Capsule

Capsules are bacterial surface structures used by many Gram-negative pathogens to evade the host immune system. They are comprised of long carbohydrate chains, called capsular polysaccharides (CPSs), which possess a lipid at one end. The lipid is connected to the CPSs through an unusual linker consisting of five to nine residues of an acidic sugar 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) (Willis and Whitfield, 2013).

In the present study, we found capsule associated genes viz. *HexA, HexB, HexC, HexD, PglA, Kmtl, KpsF, HyaE and Lctp* in various isolates of *P. multocida* under the study. Chung *et al.* (1998) cloned the entire cap locus by chromosome walking using *P. multocida* based homologous probes and found that entire cap locus spans 16 kb and contains 11 ORFs in 3 regions. Region 1 contains four ORFs, hexA, hexB, hexC and hexD (hex for haluronic acid export). Region 2 contains five ORFs, hyaA, hyaB, hyaC, hyaD and hyaE (hya for hyaluronic acid), which encode proteins that are involved in the formation of precursor activated sugar monomers and the assembly of the capsular polysaccharide polymer. Region 3 consists of two genes, *phyA* and *phyB* (phy for phospholipid substitution of hyaluronic acid) that are transcribed in the opposite direction to that of Regions 1 and 2.
HexA, hexB, hexC and hexD encode proteins responsible for transport of the polysaccharide to the bacterial surface. Gene hexA was present in P52VAC, poultry and goat isolates only. Another capsular virulence gene HexB was observed in P52VAC, poultry and buffalo isolates. For HexB, Zhang et al. (2013) compared hexB deleted mutant ΔhexB and wild-type strain P-1059 as well as the complemented strain P-1059C, which demonstrated that ΔhexB was sensitive to the bactericidal action of chicken serum, whereas P-1059 and P-1059C were both resistant. The ΔhexB was highly attenuated in chickens by intravenously injection, and intramuscular administration of ΔhexB to chickens and stimulated significant protection against P-1059 and the homologous strain X-73(A:1).

Capsule biosynthetic locus (HexC) was found in P52VAC, poultry and goat isolates and absent in cattle and buffalo, which might be due to its non coverage during sequencing run of cattle and buffalo isolates. Chung et al. (2013) cloned and sequenced entire capsule locus of avian P. multocida X-73 (A:1) and the locus was divided into three regions, region 1 of which contains four genes, hexD, hexC, hexB and hexA predicted to encode proteins responsible for transport of the polysaccharide to the bacterial surface. The sequence analysis demonstrated that the P. multocida hexABCD were highly homologous at both nucleotide and amino acid levels to Haemophilus influenza bexABCD, Actinobacillus pleuropneumoniae cpxABCD and Neisseria meningitidis ctrABCD. In the serotype A:1 strain X-73, inactivation of the capsule transport gene hexA resulted in a mutant strain that was highly attenuated in both mice and chickens, and was more sensitive to the bactericidal activity of chicken serum (Chung et al., 2001).

HexD, another capsular gene was found in P52VAC, poultry and buffalo isolates. Watt et al. (2003) demonstrated nucleotide sequence of the A:1 capsule locus
to be similar in organization to capsule loci in other bacteria, where it can be divided into three functional regions (hyaB, hexD, and phyB). Transcripts for hyaB, hexD, and phyB were detected in capsulated parental strains P-1039 and P-2100, as well as the P-2100 derivative strain (P-2100/sub), but were not detected in the non-capsulated variant strains.

*PglA*, a heparosan synthase outside of the known capsule biosynthesis region was found in all the isolates studied. *PglA* gene was previously reported in A, D and F strains of *P. multocida* only. In our experiment, it was also found in cattle, buffalo as well as in P52VAC strains which are of B type as well as in poultry isolate of A serotype and goat isolate of D serotype. Harper *et al.* (2006) reported that synthase encoded by *PglA* is transcribed 10-fold less than the synthase within the capsule operon, uses a different acceptor and gives rise to smaller molecular weight polymer products and the expression of this gene may give rise to capsular variation.

*Kmtl*, which is a *P. multocida* specific gene (for PM-PCR) was found to be present in all the isolates. This gene is routinely used for the identification of *Pasteurella* (OIE Terrestrial Manual Chapter 2.4.12 on Haemorrhagic septicemia) and all the five isolates included in the present study were confirmed before conducting whole genome sequencing by PM-PCR using the primers specific for *Kmtl* (Plate 4.2). Sequencing analysis also confirmed the presence of this gene sequence in all the five isolates. Townsend *et al.* (1998) demonstrated that hybridization of the clone KMT1, revealed binding to all serotypes of *P. multocida*; however, type B and type E isolates could be distinguished from other strains on the basis of fragment size. In addition, the clone KMT1 was able to distinguish HS-causing *P. multocida* B:2 from type B strains possessing other somatic serotypes. Arumugam *et al.* (2011) isolated 114 strains of *P. multocida* from different animal
Results and Discussion...

Species and based on the multiplex PCR assay using specific capsular genes associated with each serogroup, 114 strains were further divided to 22 species-specific *P. multocida* (KMT1 - 460 bp), 53 serogroup A (A - 1,044 bp), 33 serogroup B (B - 760 bp) and 6 serogroup D (D - 657 bp). No serogroup E (511 bp) or F (851 bp) was detected among the Malaysian *P. multocida* isolates. PCR-based typing was more discriminative and could further subtype the previously untypeable strains.

*KpsF*, arabinose 5-phosphate isomerase, was found in P52VAC, goat and cattle isolates. The *kps* locus codes biosynthesis of the capsular polysialic acid virulence factor in *Escherichia coli* K1 and contains at least two convergently transcribed operons, designated region 1 and regions 2 plus 3. Transcription through *kpsF* was required for capsule production however, complementation analysis failed to indicate a clear requirement for the *kpsF* polypeptide. *KpsF* appeared to be the first gene of a region whose gene products are required for polysialic acid transport. The function of *kpsF* is unclear, however, primary structure analysis indicated two motifs commonly observed in regulatory proteins and homology with glucosamine synthase of *Rhizobium meliloti* (Cieslewicz and Vimr, 1996). *KpsF* mediates the interconversion of ribulose 5-phosphate and arabinose 5-phosphate. *KpsF* is required for 3-deoxy-D-manno-octulosonic acid (Kdo) biosynthesis in *Neisseria meningitidis*. Mutation of *kpsF* or the gene encoding the CMP-Kdo synthetase (*kpsU*/kdsB) in *N. meningitidis* resulted in expression of a lipooligosaccharide (LOS) structure that contained only lipid A and reduced capsule expression in the five invasive disease-associated meningococcal serogroups (A, B, C, Y, and W). The step linking meningococcal capsule and LOS biosynthesis was related to Kdo production, as the expression of capsule was wild type in a Kdo transferase (*kdtA*) mutant. Thus, in
Results and Discussion...

addition to lipooligosaccharide assembly, Kdo is required for meningococcal capsular polysaccharide expression (Tzeng et al., 2002).

_HyaE_ encodes proteins required for polysaccharide biosynthesis and was found to be present only in the goat isolate. _HyaE_ shows similarity to _kjiB_, a gene in the Region 2 of the _E. coli_ K5 capsule locus. The function of this gene is not determined, but it is unique to the _E. coli_ K5 Region 2 and there is no corresponding homologue in other _E. coli_ K strains, consistent with the finding that Region 2 genes are usually capsule-type specific. The similarity of _HyaE_ to _KjiB_ is expected as the _E. coli_ K5 capsule is also a polymer containing N-acetylglucosamine and glucuronic acid with the arrangement 4-GlcA-β(1,4)-GlcNAc-α(1). The same monosaccharide components is found in the _P. multocida_ hyaluronic acid capsule. Both _hyaE_ and _KjiB_ therefore code for a function related to the common sugar monomers, in the synthesis of the N-acetylglucosamine component (Chung et al., 1998).

_LctP_, lactate permease, involved in the uptake of lactate, is present in blood and is taken up by the bacterium as a carbon energy source and also converted to precursors of capsular and lipopolysaccharide sialic acid (Exley et al., 2005). The up-regulation of _lctP_ together with the phenotype of decreased survival that was observed for the deletion mutants in _ex vivo_ model as well as in other relevant models (Exley et al., 2005) confirmed its important role and indicated that membrane transporter plays a role in increasing complement resistance of _Neisseria meningitidis_ strains. _lctP_ deletion mutant is not completely killed in human blood even after 120 minutes of incubation, which might suggest that other carbon sources could be utilized by the bacterium to generate phospho-enol pyruvate, that in turn could be used to generate sialic acid (Leighton et al., 2001).
Many virulent strains of *P. multocida* express capsule consisting of hyaluronic acid (HA) (types A and B), chondroitin (type F) or heparin (type D), which are similar to molecules found naturally in various hosts (Snipes *et al.*, 1987; Rimler and Rhoades, 1989; Tsuji and Matsumoto, 1989; DeAngelis, 1996). This 'molecular mimicry' prevents the mounting of a strong antibody response to capsular materials, impairs phagocytosis, reduces the action of complement killing and increases adherence to and survival in the host (Harmon *et al.*, 1991). Paradoxically, only *P. multocida* capsular B strains, and more specifically B:2 strains, produced hyaluronidase, an enzyme that specifically cleaves HA between its repeating disaccharides (N-acetylglucosamine and D-glucuronic acid). Decapsulation of *P. multocida* serotype A:3 with hyaluronidase increased phagocytosis by macrophages, whereas an unencapsulated variant of the bacteria was not internalized (Pruimboom *et al.*, 1996).

### 4.3.2.1.2 Fimbriae and adhesion

Microbial infections are typically initiated by the colonization of tissues by a specific mechanism that promotes adherence to host cells or tissues. A necessary step for the development of an infectious disease in the host organism is the formation of a firm link between the pathogen and the target cells of the host. The link is mediated by surface-exposed adhesive organelles and is required for internalization of bacteria or extracellular colonization of host tissues. The adhesive organelles mediate bacterial adhesion via a specific interaction with surface structures present on host cells. The adhesin binding to the target cells triggers subversive signals that allow pathogens to evade immune defense and facilitate bacterial colonization or invasion (Zav’yalov *et al.*, 2010). Fimbriae play a role in the surface adhesion, as fimbriae on some *P. multocida* serotype A strains that were able to adhere to mucosal epithelium, but not
on the surface of those strains unable to adhere (Glorioso et al., 1982; Rebers et al., 1988; Isaacson and Trigo, 1995; Ruffolo et al., 1997). Type IV fimbriae (pili) from *P. multocida* serotypes A, B and D (Ruffolo et al., 1997) are associated with virulence in other bacteria because of their role in attachment to host cell surfaces. However, the role of fimbrial structures in *P. multocida* virulence is still unproven (Harper et al., 2006).

ComE is a fibronectin binding protein which is non fimbrial adhesion protein identified as a virulence factor in *P. multocida*. Mullen et al. (2008) identified ComEl (12.7 kDa) using phage library of *P. multocida* and shown to interact with integrin-binding fibronectin type III (FnIII) repeats and to be involved in bacterial competence and DNA uptake (Shivachandra et al., 2011). This protein shares many structural and functional similarities with the fibronectin binding proteins of *S. aureus* and *S. pyogenes*, suggesting that this is the general mechanism of bacterial binding to fibronectin. In our study also, ComE was found to be present in all the five isolates except for P52VAC strain of *P. multocida*.

ClpB, caseinolytic peptidase B is a ATP-dependent chaperone (Heat shock protein) was present in poultry, goat and cattle isolates. ClpB belongs to a subclass of highly conserved proteins that include molecular chaperones, proteolysis regulators, and regulators of thermotolerance. It plays an important role in protection or recovery from a variety of stresses. In *E. coli*, clpB mutants exhibit a higher death rate at high temperatures, underscoring the importance of ClpB in thermotolerance. In *Saccharomyces cerevisiae*, the ClpB homolog, *Hsp104*, is also important for recovery from thermal stress as well as a variety of other stressful conditions (Fuge and Farr, 1993). Nanduri et al. (2009) studied the effects of three classes of antibiotics viz. amoxicillin, chlortetracycline, and enrofloxacin on the *P. multocida* transcriptome.
using custom oligonucleotide microarrays and identified ClpB as a potential virulence factor which was down-regulated leading to decreased virulence.

HofC, an assembly protein in type IV pilin biogenesis, was found to be present in all the five isolates except for the cattle. Voss et al. (2014) reported that most Hop and Hom OMPs were susceptible to proteolysis, whereas Hor and Hof proteins were relatively resistant to extracellular protease digestion in Helicobacter pylori. Protease-susceptible OMPs contain a large protease-susceptible extracellular domain exported beyond the outer membrane and a protease-resistant domain at the C-terminus with a predicted β-barrel structure. Their results provided insights into the repertoire of surface-exposed H. pylori proteins, which may mediate bacterium-host interactions as well as the cell surface topology of these proteins.

Hsf is a fimbriae gene and was present in all the isolates of P. multocida studied. Fimbriae play a role in surface adhesion and attachment to host cells and extracellular matrix protein is a primary prerequisite for bacterial adhesion. This implies role of fimbriae as a potential virulence factor required for bacteria to induce disease process after entry in the host.

PfhB1 and PfhB2 are filamentous hemagglutinin proteins, and were present in all the five isolates except for the cattle, in which pfhB1 was absent. P. multocida possesses two filamentous hemagglutinins, PfhB1 and PfhB2 (also named FhB1 and FhB2, respectively). P. multocida Pm70 genome sequence predicted two genes encoding for the filamentous hemagglutinin proteins (FhaB1 and FhaB2) that appear to increase bacterial dispersion and colonization (Fuller et al., 2000; May et al., 2001). These proteins are similar to the LspA1 and LspA2 filamentous hemagglutinins from Haemophilus ducreyi, mutations of which have been shown to affect virulence. A similar filamentous hemagglutinin in Bordetella pertussis and B.
bronchiseptica is required for biofilm formation and colonization of the nose and trachea in mice. FhaB2 in *P. multocida* has been implicated in virulence, and expression of FhaB2 was found to be reduced by 4-fold in a non-mucoid *P. multocida* variant, AL1114. FhaB and its transporter FhaC form a two-partner secretion system that is similar to the FhaB-FhaC system in *Bordetella* species, LspA-LspB in *H. ducreyi*, and IbpA-IbpB in the pathogenic *Histophilus somni* strain 2336, which act to inhibit phagocytosis (Wilson and Ho, 2013). These genes share significant similarity with a class of genes that encode filamentous haemagglutinins, which in *Bordetella pertussis* plays a major role in the colonization of the upper respiratory tract (Kimura *et al.*, 1990; Mooi *et al.*, 1992). Mutation of these genes in *P. multocida* resulted in significantly reduced virulence in mice (Fuller *et al.*, 2000) and pfhA2 mutant strain P1059 was shown to be highly attenuated in turkeys when administered intranasally, but only moderately attenuated when given intravenously (Tatum *et al.*, 2005). Ewers *et al.* (2006) showed that only 46% of the cattle isolates investigated harbored the *fha* gene, and a significant association was reported between disease status and bovine *P. multocida* strains harboring *pfhA* gene. Hypothetically, *P. multocida* *fha* may also function as a serum resistance factor based on sequence homology to *H. somni* immunoglobulin binding p76 protein (Cole *et al.*, 1993).

PfhR, hemoglobin receptor and iron transport, was present in all the isolates of *P. multocida*. Wheeler (2009) separated OMPs (Outer membrane proteins) by SDS-PAGE and identified by matrix-assisted light desorption/ionization time of flight mass (MALDI-TOF) in the high molecular weight (HMW) region. Iron-acquisition OMPs expressed by isolates were OMPs HasR, HemR, PfhR and hypothetical protein PM0741 from the invasive disease group as well as in porcine and ovine pneumonic group. HemR and PfhR were common to all invasive disease isolates. OVine isolate
had a very similar OMP profile to porcine in iron-rich culture, and had reduced HMW OMP expression. Similar gene was present in the complete sequence of avian and porcine isolates of *P. multocida* (E-komon *et al.*, 2012).

*PlpB*, outer membrane lipoprotein 2, was found to be present in all the isolates studied. This gene encodes 39 kDa cross-protecting protein of *P. multocida* (Tabatabai and Zehr, 2004) and is predicted to be an ABC transport protein required for the uptake of methionine into the cell (Merlin *et al.*, 2002). Lipoprotein B (*PlpB*) from *P. multocida* serotype A isolate was shown to be an adherence factor which stimulated cross-protective immunity in mice and poultry (Dabo *et al.*, 2008). Expression of *PlpB* was related to the presence and amount of capsule present on the cell (Ali *et al.*, 2004b). This gene is essential for adherence to the host cell and was present in all the isolates of *P. multocida*. Wheeler (2009) also reported presence of *PlpB* in iron limited culture. Cooney and Lo (1993) reported the presence of *PlpB* in avian isolates only, whereas in our study, it was in addition to the avian isolate, also found in cattle, buffalo and goat isolates.

*Pasteurella* lipoprotein E (*PlpE*) is a protective surface antigen associated with *P. multocida* serotype A:1, A:3 and A:4 strains isolated from cattle with shipping fever (Wu *et al.*, 2007) and from birds with fowl cholera (Hatfaludi *et al.*, 2012). *PlpE* gene was found to be present in P52VAC, poultry and goat isolates in our study, which is consistent with the study of Wu *et al.* (2007) and Hatfaludi *et al.* (2012), as *PlpE* gene was found to be present in P52VAC which is a serotype B:2 strain. Using a bioinformatics approach, 98 genes in avian strain Pm70 and 107 genes in the nontoxinogenic porcine strain 3480 were identified as encoding putative OMPs (E-komon *et al.*, 2012). Of this combined list, 71 recombinant proteins were expressed and purified, albeit most as insoluble proteins, and tested as vaccine candidates. Only one
protein, lipoprotein E (PlpE), was found to protect against *P. multocida* challenge in chicken and mice, which confirmed previously reported results using the PlpE cloned from the avian serotype A:1 strain X-73 (Wu et al., 2007). However, a PlpE knockout mutant strain retained full virulence (Hatfaludi et al., 2012). Conjugated vaccines comprised of multiple antigens, such as OmpH plus PlpE peptides, have also shown promise (Okay et al., 2012a; Okay et al., 2012b). Pandher et al. (1998) cloned and sequenced the gene encoding protein PlpE which is similar to an *Actinobacillus pleuropneumoniae* lipoprotein, OmL. Affinity-purified, anti-PlpE antibodies recognized a protein in all serotypes of *P. haemolytica*. *P. haemolytica* and recombinant *E. coli* expressing PlpE are capable of absorbing anti-PlpE antibodies from bovine immune serum, which indicated that PlpE is surface exposed in *P. haemolytica*. In complement-mediated killing assays, a significant reduction in killing of *P. haemolytica* was observed by them when bovine immune serum was depleted of anti-PlpE antibodies, and used as the source of antibody. Results indicated PlpE is surface exposed and immunogenic in cattle and that antibodies against PlpE contribute to host defense against *P. haemolytica*. In our experiment, PlpE was found to be present in P52VAC, poultry and goat isolates. Similarly, Abrahante et al. (2013) also found PlpE gene coding for a cross-protection factor antigen present in all three strains (two virulent avian strains and Pm70) and highly conserved.

PlpP, involved in methionine binding, was found to be present in all the isolates of *P. multocida* studied except the cattle isolate. Initially, recombinant PlpB failed to protect either mice or chicken against *P. multocida* challenge but later it was found that PlpE is cross-protective (Wu et al., 2007). Immunisation with recombinant *P. multocida* PlpE conferred protective immunity, with 80-100% of mice and 63-100% of chicken were protected against heterologous challenge. PlpE is the first *P.
*multocida* recombinant protein to stimulate high level cross-serotype protective immunity.

RcpA (Rough colony protein A) and RcpB (Rough colony protein B), are outer membrane secretin family protein components of Tad (Tight adherence), which were found to be absent only in buffalo isolate. Tight adherence (*Tad*) locus, comprised of genes encoding for outer membrane proteins (OMPs) (*TadD, RcpA, RcpB* and *Flp1*), is involved in the Tad macromolecular transport system that is required for biofilm formation and colonization (May *et al.*, 2001). RcpA forms the multimeric outer membrane secretion channel (secretin) of the Flp pilus biogenesis apparatus. The 43-kDa protein (RcpA) was similar to precursor protein D of the general secretion pathway of Gram-negative bacilli, while the 20-kDa protein (RcpB) appeared to be unique. The genes encoding these proteins have been cloned from *Actinobacillus actinomycetemcomitans* 283 and sequenced. A BLASTX (gapped BLAST) search of the surrounding ORFs revealed homology with other fimbria-related proteins. Haase *et al.* (1999) reported that the genes encoding the 43-kDa (*rcpA*) and 20-kDa (*rcpB*) proteins may be functionally related to each other and to genes that may encode fimbria-associated proteins. The genes *TadB* and *RcpA* are part of a large operon which, in *A. actinomycetemcomitans*, is composed of 14 genes and mediates nonspecific adhesion to solid surfaces, whether they are biological surfaces or not (Auger *et al.*, 2009). *TadB* is a part of *flp* (fimbria like protein) operon, a widespread locus in bacteria known to be essential for virulence in *A. actinomycetemcomitans* and *Haemophilus ducreyi*. The expression of *P. multocida* *tad* genes was also up-regulated under starvation conditions (Paustian *et al.*, 2001).

The *flp-1* gene is found upstream of the *tad* operon in a low G+C region (similar G+C ratio to the *tad* operon) along with two other genes (rcpA and rcpB)
expressed specifically in rough adherent bacteria but not in smooth mutants (Haase et al., 1999). The relationship between the tad operon and flp-1 is that the former is a secretion system for the later.

*TadE* and *TadF* gene loci code for tide adherence (Tad), which was found to be absent in buffalo isolate (for *TadE*) and in cattle as well as buffalo isolates (for *TadF*). The tad locus of *Actinobacillus actinomycetemcomitans* encodes a molecular transport system required for tenacious, non-specific adherence to surfaces and formation of extremely strong biofilms. This locus is dedicated to the biogenesis of Flp pili, which are required for colonization and virulence and 11 of the 14 *tad* locus genes are required for adherence and Flp pilus production (Perez et al., 2006). *TadE* and *TadF* are pilin-like proteins, which are termed as pseudopilins (Tomich et al., 2006). Using site-directed mutagenesis, Tomich et al. (2006) demonstrated that for processing of pre-Flp1, pre-*TadE* and pre-*TadF* are required for biofilm formation.

The LtxA toxin of *A. actinomycetemcomitans* had been thought to be entirely cell associated; either bound to cell surface-associated nucleic acids or within membranous vesicles which bud from the bacterium’s surface. This means that the bacterium itself is toxic to target cells. Detailed analysis has revealed that adherent (or rough) colonies of *A. actinomycetemcomitans* do not secrete leukotoxin but non-adherent (smooth colonies) do excrete leukotoxin. Bacteria with mutations in the Tad operon were found to secrete the leukotoxin, suggesting that binding of this toxin to *A. actinomycetemcomitans* is in some manner dependent on proteins expressed by *Tad* genes.

*Plp4*, type IV fimbriae (pili), was found to be present in all the studied isolates of *P. multocida*. Pilli are associated with attachment of the bacteria to host cell surfaces in other bacteria also. Type IV fimbriae (pili) have been isolated and
characterized from *P. multocida* serotypes A, B and D (Ruffolo *et al.*, 1997) but not from cattle and buffalo. Cloning and sequence analysis of the ptfA gene from *P. multocida* B:2 (strain p52) indicated 78.4% similarity with that of the A:1 strain (Siju *et al.*, 2007). Treatment of the organism with pronase, heat, acidity or homogenization confirmed the involvement of the fimbriae in the attachment process. An anti-fimbrial antibodies specifically inhibited the adherence capabilities of different serotypes that are known to greatly vary for tracheal epithelial cells from different hosts (Harper *et al.*, 2006). This might be the first report of type IV fimbriae in *P. multocida* of cattle, buffalo and goat.

4.3.2.1.3 Iron metabolism

Virtually all bacterial pathogens require iron to infect vertebrates. The most abundant source of iron within vertebrates is in the form of heme as a cofactor of hemoproteins. Many bacterial pathogens have elegant systems dedicated to the acquisition of heme from host hemoproteins. Once internalized, heme is either degraded to release free iron or used intact as a cofactor in catalases, cytochromes, and other bacterial hemoproteins. Paradoxically, the high redox potential of heme makes it a liability, as heme is toxic at high concentrations. Although, a variety of mechanisms have been proposed to explain heme toxicity, the mechanisms by which heme kills bacteria are not well understood. Bacteria employ various strategies to protect against and eliminate heme toxicity. Factors involved in heme acquisition and detoxification have been found to contribute to virulence, underscoring the physiological relevance of heme stress during pathogenesis (Anzaldi and Skaar, 2010).

Iron homeostasis in the mammalian host limits the availability of iron to invading pathogens and is thought to restrict iron availability for microbes inhabiting
mucosal surfaces. The presence of surface receptors for the host iron-binding glycoproteins transferrin (Tf) and lactoferrin (Lf) in globally important Gram-negative bacterial pathogens of humans and food production animals suggested that Tf and Lf are important sources of iron in the upper respiratory or genitourinary tracts, where they exclusively reside (Morgenthau et al., 2013).

ExbB, a biopolymer transport protein, functions as a part of the TonB-dependent energy transduction system for the import of iron-siderophore complexes and vitamin B12 across the outer membrane. The iron transport protein energy-transducing genes exbB, exbD and tonB were expressed at 2.1-fold elevated levels when transferrin was added to iron-free medium. Some of the other genes that were elevated in expression included those encoding the tryptophan transport protein Mtr, the transmembrane permease FbpB, and the branched-chain amino acid transporter AzlD. In the present study, ExbB gene was found to be present in all the isolates sequenced suggesting the importance of this gene in virulence of P. multocida.

The ferric iron-binding protein (Fbp) expressed by pathogenic Neisseria spp. plays a central role in the high-affinity acquisition of iron from human transferrin. Fbp participates in this process as a functional analogue of a Gram-negative periplasmic-binding protein component, which operates as a part of a general active transport process for the receptor-mediated, high-affinity transport of iron from human transferrin (Chen et al., 1993). FbpA functions in complexing and transporting iron across the periplasmic space as the periplasmic component of the iron ABC transport system. The ABC family of transporters includes a broad and diverse group of import systems found in prokaryotes. ABC transporters generally consist of separate polypeptides that form an A1B2C2 complex at the inner membrane of the cell during transport. Thus, in order for the ferric ion to reach the cytoplasm, it is donated to an
inner membrane complex consisting of the inner transmembrane protein FbpB and the cytoplasmic ATPase FbpC. This high affinity bacterial iron uptake system results in iron from host transferrin being liberated at the cell surface, translocated through the periplasmic space, and deposited into the cytoplasm for use or storage without the incorporation of the transferrin protein into the bacterial cell (Shouldice et al., 2003).

Because the FbpABC pathway appears to be a common route for ferric iron uptake from several different sources, it is particularly an attractive therapeutic target. In our study also, we found the presence of FbpA, FbpB and FbpC in all the five isolates except for absence of FbpA and FbpC in cattle and buffalo isolates, respectively. This indicated importance of this FbpA, FbpB and FbpC gene cluster for the virulence or disease progression by P. multocida.

HbpA, a haemin-binding protein is found to be regulated by iron, manganese and haemin through a Fur-independent mechanism. One protein implicated in heme utilization is the heme-binding lipoprotein (HbpA). HbpA was initially identified as a potential constituent of a heme acquisition pathway following transformation of H. influenzae genomic DNA library into Escherichia coli and screening for recombinant clones with heme-binding activity. Expression of heme-binding activity by E. coli correlated with the expression of a protein of approximately 51-kDa, sized on SDS-PAGE gels, that was subsequently purified in a heme-agarose affinity purification protocol, from both recombinant E. coli and H. influenzae, and shown to be a lipoprotein. HbpA was localized to the periplasmic space and shown to be associated with both the inner membrane and the outer membrane in H. influenzae. HbpA serves to transport heme into the cytosol of H. influenzae subsequent to initial binding steps at the cell surface. However, no direct evidence has been presented for the role of HbpA in heme utilization (Morton et al., 2005).
Results and Discussion...

Garrido et al. (2003) inoculated purified HbpA protein into Swiss mice previous to a challenge treatment with wild-type *P. multocida* cells. No protective effect was obtained as compared to whole outer membrane proteins of this microorganism used in the immunization procedure. This indicated that either other haemin-binding proteins are present in *P. multocida* cells or blocking HbpA by antibodies is not sufficient to prevent the infective process of the mice cells presumably because other iron receptors present in this bacterium are able to acquire the necessary amount of iron for survival.

HbpA, a haemin binding protein in immunization assays is not able to protect mice against a challenge with virulent *P. multocida* cells. These data are in agreement with the existence in *P. multocida* of several putative haem- or haemoglobin binding proteins (May et al., 2001). HbpA was found to be present in all the isolates of *P. multocida* in our study. HbpA gene needs to be fully characterized for its role in disease due to *P. multocida*, as it may not be directly involved in virulence but it may work as virulence associated genes or protein.

HgbA is a necessary surface-located OM protein of bacterial species that interacts with Hb to import haem into the cell. HgbA may be sufficient as sole Hb receptor or HgbA may function in concert with another protein to mediate haem import. This later mechanism resembles transferrin mediated iron uptake in which two OM proteins, TbpA and TbpB, are required for efficient iron uptake (Srikumar et al., 2004). Fuller et al. (2000) identified two loci containing potential virulence genes with homology to iron-acquisition-related genes *ExbB* and *HgbA* of *H. influenzae*. Bosch et al. (2002b) demonstrated that three ORFs (PM0298, PM0299 and PM300) of *P. multocida* constitute a single transcriptional operon. Mutants defective in PM0298 or PM0299, demonstrated that these genes were essential for the viability of the
pathogen (Bosch et al., 2002b) The HgbA (PM300) gene product bound hemoglobin (Bosch et al., 2002b) and hemin (Bosch et al., 2004), but binding and virulence of the hgbA-mutant strain were not affected (Bosch et al., 2002b). The hgbA gene is widespread in P. multocida strains regardless of the serotype or the source of the isolate. So, hgbA gene may be useful as a specific probe to detect P. multocida in field samples. In our study also, it was found to be present in all the five isolates studied. It has been demonstrated that some TonB-dependent iron receptors, such as transferrin and lactoferrin binding proteins, are able to induce protection when they are used as prophylactic immunogens. HgbA protein may also be useful as an antigen that provides protection against virulent P. multocida strains (Bosch et al., 2002a).

The prevalence of tonB, tbpA, hgbA and hgbB genes in P. multocida bovine strains is reported to be 100, 70.2, 95.2 and 57.7%, respectively (Ewers et al., 2006), and there was significant association between disease status and the presence of hgbB and tbpA as single virulence associated genes in P. multocida bovine isolates. The potential for TbpA to be a good vaccine candidate is questionable, because approximately 30% of bovine P. multocida strains lack the TbpA gene (Ewers et al., 2006; Roehrig et al., 2007).

HemR protein is involved in binding hemopexin bound heme from serum important for iron and porphyrin acquisition by a number of pathogens. HemR expressing pathogens enable the pathogen to obtain iron from a large number of heme containing proteins including hemoglobin, myoglobin, hemealbumin, and catalase, ensuring survival in the host (Tabatabai, 2008). HemR protein may be expressed only during hemin-limiting conditions (the environment of the human oral cavity). Since the HemR protein has significant homology with other iron-regulated outer membrane proteins, a potential role for this HemR protein in hemin transport was suggested
Results and Discussion...

(Karunakaran et al., 1997). HemR was found in all the isolates included in our study except for P52VAC.

HemU is a transmembrane protein required for transfer of hemin across the membrane, and was found to be present in all the isolates of our study. Heme, a major iron source, is transported through the outer membrane of Gram-negative bacteria by specific heme/hemoprotein receptors and through the inner membrane by heme-specific, periplasmic, binding protein-dependent, ATP-binding cassette permeases. Yersinia enterocolitica strains mutated in \textit{hemU} or \textit{hemV} genes were unable to use haemin as an iron source whereas those mutated in the \textit{hemT} gene were able to use haemin as an iron source. Escherichia coli strains expressing only the haemin outer membrane receptor protein HemR from \textit{Y. enterocolitica} were capable of using haemin as an iron source.

OmpW, an outer membrane protein is involved in protective immunity in \textit{P. multocida} infections and the OMPs expressed exclusively \textit{in vivo} may be involved in stimulating immunity to serotypes expressing different LPS types. It is also likely that many \textit{P. multocida} virulence factors will be surface located, and these may include factors critical for colonisation and invasion, important steps in pathogenesis which are poorly understood (Boyce et al., 2006). In our study, OmpW was found to be present in all the isolates imparting their role and importance in the pathogenesis of \textit{P. multocida}. Further studies need to be carried out for the complete characterization of OmpW protein in the diseases caused by \textit{Pasteurella}.

Rfb is a capsule transport protein and the O antigen is encoded by the \textit{rfb} gene cluster. The Rfb gene was also present in all the isolates of \textit{P. multocida} in our study. The Rfb region contains all the information necessary for synthesis of the components of the O-antigen, its assembly and export to the cell surface as demonstrated by
Results and Discussion...

expression in a heterologous host (Morona et al., 1991). Various rfb mutations have effects on the rhamnose contents of the cell wall polysaccharides produced, including a loss of O antigen production, a reduced level of LPS production, or production of LPS with a reduced amount or complete lack of rhamnose (Boels et al., 2004). This gene also plays an important role by decreasing virulent component of the bacteria and thus imparts its role in pathogenesis of pasteurellosis.

*RffG* codes for the enzyme dTDP-D-glucose-4,6-dehydratase and participates in formation of O-specific polysaccharide or O antigen, which, joined together with lipid A via core oligosaccharide, forms lipopolysaccharide in the bacterial outer membrane. The enzyme also participates in formation of the polysaccharide part of the enterobacterial common antigen, a cell surface glycolipid. Gene *RffG* was present in all the five isolates of *P. multocida* in the present study and not fully characterized for their role in bacterial virulence and pathogenesis.

*TbpA* proteins are TonB-dependent, OM receptors that are thought to serve as channels through which ferric iron crosses the outer membrane after its release from bound transferrin (Ratledge and Dover, 2000). *TbpA*, transferin binding protein, was found to be present in all the isolates of *P. multocida* except for the goat isolate. *TbpA* binds to bovine transferrin and bovine *P. multocida* are able to utilize iron from bovine transferrin, but not from other species. In constrast, avian *P. multocida* are unable to utilize iron from any transferrin molecules including from avian sources (Hatfaludi et al., 2010). Ewers et al. (2006) was not able to amplify *TbpA* gene from avian and pig strain which is in contrast with our result as *TbpA* was found to be present in the poultry and bovine isolates and absent in the goat isolate. Absence of *TbpB* gene in all our isolates is consistent with previous studies stating none of their isolates having *TbpB* gene in their genome (Wheeler, 2009). *TbpA* from *P. multocida*...
serotype B:2 strain was predicted to be a gated-pore with several membrane-spanning regions (Shivachandra et al., 2005). The specific role of TbpA in *P. multocida* virulence is still unknown.

TonB is a periplasmic protein and was found to be present in all the isolates studied except for the cattle. Transport of the iron into the bacterial cell by any of high-affinity systems requires the product of the *tonB* gene, which links both cytoplasmic and outer membranes thus enabling the transfer of energy necessary for this process involving tonB-dependent lactoferrin and transferrin receptor (Bosch et al., 2002a). The function of TonB is apparently to transduce energy to the outer membrane receptors, by releasing tightly bound ligands into the periplasmic space. The ExbB and ExbD proteins, which are in the inner membrane stabilizing the TonB protein was present in all the isolates of poultry studied by them, which signifies its importance in iron utilization. A conserved region present in these receptors, known as the TonB box, is responsible for the interaction between them and the TonB protein. Binding of the iron-loaded molecules (siderophore, transferrin, etc.) to the outer membrane receptor induces a conformational change in it, which enables the iron to pass through the TonB pore (Ratledge and Dover, 2000). A TonB mutation reduces virulence in *Haemophilus influenzae*, *Bordetella pertussis* and *Pseudomonas aeruginosa*. Likewise, a wild-type TonB gene is required for intracellular growth and the intercellular spread of *Shigella dysenteriae*. ExbB, ExbD and TonB mutants of *P. multocida* showed poor growth in both BHI plates and liquid medium. All three components of ExbB, ExbD and TonB are equally required to support the infectious process of *P. multocida* cells (Bosch et al., 2002a).

TonB-dependent lactoferrin and transferrin receptor was absent only in the goat isolate whereas, it was present in bovine, bubaline and avian isolates in our
study. Transferrin (Tf) is the predominant iron-containing extracellular protein within the body; therefore, Tf receptors play a critical role during infection. Lectoferrin (Lf) receptors may play an important role during colonization, but the relative importance for iron acquisition and protection from cationic antimicrobial peptides is uncertain. Assumptions that Lf receptors were primarily important on mucosal surfaces need to be revisited and the relative role in iron acquisition and protection from cationic antimicrobial peptides needs to be assessed (Morgenthau et al., 2013).

TonB-dependent receptor was found to be present in poultry and goat isolates only, whereas it was absent in P52VAC, buffalo and cattle isolates or so as to say in bovine isolates. Ogunnariwo and Schryvers (2001) analyzed bovine respiratory isolates of *P. multocida* and demonstrated that six of nine strains tested were capable of growth dependent upon bovine transferrin and of specifically binding ruminant transferrins. In contrast to other species, binding of this protein to immobilized transferrin was specifically blocked by the N-lobe sub-fragment of bovine transferrin. *P. multocida* has a single, novel receptor protein (TbpA), that is capable of efficiently mediating iron acquisition from bovine transferrin without the involvement of a second receptor protein (TbpB).

Translocation protein TolB, is a periplasmic protein of the cell envelope Tol complex. It is partially membrane associated through an interaction with the outer membrane lipoprotein AL (peptidoglycan-associated lipoprotein), which also belongs to the Tol system. The interaction of TolB with outer membrane porins of *Escherichia coli* was investigated with a purified TolB derivative harboring a six-histidine tag. TolB interacted with the trimeric porins OmpF, OmpC, PhoE, and LamB but not with their denatured monomeric forms or OmpA. These interactions took place both in the presence and in the absence of lipopolysaccharide. Interactions of TolB with porins
might take place \textit{in vivo} and participate in porin assembly. The Tol system as a whole may be involved in porin assembly in the outer membrane (Rigal \textit{et al.}, 1997). TolB is a periplasmic protein that peripherally associates with the outer membrane and participates in peptidoglycan metabolism or OM lipoprotein assembly. TolB proteins are involved in maintaining cell envelope integrity of \textit{Escherichia coli}. TolB is found to be periplasmic, although it is partially membrane-associated and interact with membrane proteins (Bouveret \textit{et al.}, 1995). TolB box is found in various TolB-dependent colicins and is required for the interaction of colicin with TolB. Colicins are proteins produced by some strains of \textit{Escherichia coli} that are lethal for related strains of \textit{E. coli}. In our study, TolB protein was found to be present in P52VAC strain only, which is generally used for the vaccination of animals against HS disease.

4.3.2.1.4 Outer membrane proteins

Outer membrane (OM) functions as a selective barrier that prevents the entry of many toxic molecules into the cell, a property that is crucial for bacterial survival in many environments. At the same time, the proteins embedded in the OM fulfill a number of roles that are critical for the bacterial cell, such as nutrient uptake, transport of molecules in and out of the cell, and interaction with the environment and host tissues. As cytoplasmic proteins are usually soluble and easier to characterise than membrane proteins, they have often been the subject of proteomic studies. Outer membrane protein (OMP) of Gram negative bacteria has a role in disease processes as it acts at an interface between the host and pathogen (Lin \textit{et al.}, 2002). Thus, OMP variation among the isolates may help in epidemiological survey by assessing their inter-strain heterogeneity and can be used to assess intra-species diversity (Davies \textit{et al.}, 2003). Despite progress in recent years, detailed knowledge of membrane proteins remains elusive (Hatfaludi \textit{et al.}, 2010).
HasR, a heme acquisition system receptor in a bovine *P. multocida* isolate, is a protein, which is 98% identical to its homolog in the *P. multocida* Pm70 genome sequence, is surface exposed, is conserved among most *P. multocida* isolates and is an immunodominant iron-regulated outer membrane proteins (IROMPs) (Prado *et al.*, 2005). It is expressed under low-iron conditions *in vivo* and confers protection against challenge with ovine *P. multocida* serotype A:3 strain 232. This indicated that antibody responses in cattle are induced by *P. multocida* IROMPs, and that the 96 kDa HasR protein is an immunodominant IROMPs (Prado *et al.*, 2005). A whole-cell vaccine based on a serotype A:1 strain of *P. multocida* that has been inactivated by treatment with high iron concentrations has also been explored (Herath *et al.*, 2010). HasR protein has been described for *P. multocida* A:3 in cattle as an immunodominant, presumably a *fur* (ferric uptake regulator) gene dependent iron-regulated outer-membrane protein as is the expression of the HemR protein (Tabatabai, 2008). Dabo *et al.* (2008) reported that vaccination of calves with OMPs or IROMPs resulted in significant decrease in lesion scores following *P. multocida* challenge compared to vaccination with adjuvant alone. Vaccinated calves developed intense antibody responses to a 96 kDa protein band, and correlation between high anti-96 kDa antibodies and low lesion scores approached significance (P<0.06). The 96 kDa OMP was a homologue to the iron-regulated protein HasR, a heme acquisition receptor protein. In our study, HasR was found to be present in all the isolates of *P. multocida*.

LppB codes for lipoprotein B protein, and contains a prokaryotic membrane lipoprotein lipid attachment site profile from amino acids 1 to 17, and a LysM domain from amino acids 120 to 163 that is also found in a variety of enzymes involved in bacterial cell wall degradation and have a general peptidoglycan binding function. An
experimental vaccine enriched with these recombinant lipoproteins generated high antibody titers in rabbits and sheep and exerted a protective effect in mice against septicemia induced by *Histophilus somni* bacterial challenge (Guzman-Brambila et al., 2012). The outer-membrane lipoprotein-encoding gene, *lpp* was the most highly expressed gene during growth of *Y. pestis* in human plasma. A proteome analysis showed that Lpp is highly immunogenic upon *Y. pestis* EV76 injection into rabbits (Li et al., 2005). Lpp represents a major constituent of the *Y. pestis* membrane, and may be important at various steps of the disease. Braun (murein) lipoprotein (Lpp), a major bacterial outer membrane component of Gram-negative bacteria in the family *Enterobacteriaceae*, contributed significantly to the development of septic shock. Virulence of *Salmonella enterica* is abolished by mutating one or the other of the two copies of the *Lpp* gene (Sha et al., 2004) was also reported. In the present study, *LppB* gene was present in all the isolates of *P. multocida*.

LspB, a hemolysin secretion and activation protein, was found to be present in all the isolates of *P. multocida* studied. Haem-uptake systems are associated with hemolysins that liberate hem from host erythrocytes and other cell types. *P. multocida* is described as non-hemolytic under aerobic conditions. However, studies have reported production of an inner membrane-bound hemolysin, expressed under anaerobic conditions in vitro and is most active against avian red blood cells. The hemolysin production could enhance hem acquisition under certain in vivo conditions and could be associated with increased virulence of strains (Wheeler, 2009).

OmpH, Protein H, or porin H, is the major outer membrane protein in the envelope of *P. multocida*. This protein has been purified and characterized as a porin because it is structurally and functionally related to the superfamily of porins of
Gram-negative bacteria. Bacterial porins are channel-forming transmembrane proteins which are found in the outer membranes of Gram-negative bacteria. They function as molecular sieves to allow the diffusion of small hydrophilic solutes through the outer membrane and also serve as receptors for bacteriophages and bacteriocins. Porins are highly immunogenic, exposing epitopes on the bacterial surface. They are generally conserved in a bacterial species or even in a bacterial family in that they have high homology in primary amino acid sequence and secondary structure and are antigenically related. These properties make porins attractive vaccine candidates for induction of homologous and heterologous immunity against Gram-negative bacterial infections. In native conformation, porin H is a homotrimer, stable in sodium dodecyl sulfate (SDS) at room temperature, and is dissociated into monomers upon boiling. The molecular masses of denatured monomers range between 34 and 42 kDa depending on the serotype and the electrophoretic system used for analysis (Luo et al., 1997). In our study, we found the presence of OmpH in all the isolates studied.

Outer membrane lipoprotein LolB, is an essential lipoprotein that functions as an outer membrane receptor for lipoproteins. LolB is one of the essential outer membrane lipoproteins, being involved in the last step of lipoprotein sorting. It accepts lipoproteins from a periplasmic molecular chaperone, LolA, and mediates the outer membrane anchoring of lipoproteins through a largely unknown mechanism (Tsukahara et al., 2009). The release and outer membrane localization of LolB per se are also dependent on LolA and LolB, respectively. All Lol proteins are essential for E. coli growth (Tokudaa and Matsuyama, 2004). Outer membrane lipoproteins of Escherichia coli are released from the inner membrane upon the formation of a complex with a periplasmic chaperone, LolA, followed by localization to the outer membrane. Localization of lipoproteins to the outer membrane generally requires an
Results and Discussion...

outer membrane lipoprotein, LolB, and occurs via transient formation of a LolB-lipoprotein complex. A mutant carrying the chromosomal LolB gene under the control of the lac promoter-operator grew normally in the absence of LolB induction, if the mutant did not possess the major outer membrane lipoprotein Lpp, suggesting that LolB is only important for the localization of Lpp. For in vivo function of LolB, Tanaka et al. (2001) constructed a chromosomal LolB null mutant harboring a temperature-sensitive helper plasmid carrying the LolB gene. At a nonpermissive temperature, depletion of the LolB protein due to loss of the LolB gene caused cessation of growth and a decrease in the number of viable cells (Tanaka et al., 2001).

In our study, LolB was found in P52VAC and cattle isolates only.

Filamentous haemagglutinin outer membrane protein, pfhA, showed a higher prevalence in capsular types A and F compared with capsular type D, and the HghB gene was more frequently detected in capsular D isolates than in capsular A isolates (Ewers et al., 2006; Tang et al., 2009). Gene pfhA was also not found in any of the four field isolates analyzed in our study. Ewers et al. (2006) also found this gene in 25% (2/8) of isolates from dogs, 18.5% from cats (10/54), and 21.2% (11/52) from swine. Tang et al. (2009) working with swine strains described that 15% of P. multocida tested positive to the pfhA gene (35/233). In our study also, we found the presence of filamentous hemagglutinin outer membrane protein in only P52VAC.

PtfA gene products assemble to form type IV fimbriae on the bacterial surface. This subunit is exported from the cell and polymerised to form the fimbrial strand by a complex multi-component pathway. Doughty et al. (2000) isolated type 4 fimbrial subunit protein (PtfA) from whole membrane fraction and identified as an 18-kDa protein. Sellyei et al. (2010) reported that among fimA (fimbriae), hsf-1,2 (autotransporter adhesins), pfhA (filamentous hemagglutinin), tad (nonspecific tight
adherence protein) and ptfA (subunit of type 4 fimbriae), only fimA, hsf-2 and ptfA are present in all or virtually all pathogenic isolates of P. multocida. These indicated a possible role of ptfA gene in the virulence of P. multocida in various hosts and disease conditions. Ewers et al. (2006) detected ptfA (99.0%), coding for a type IV fimbrial subunit as well as the outer membrane protein encoding gene oma87 (99.7%) in most of the strains. The PtfA gene was demonstrated in most bovine P. multocida isolates and has been cloned (Adler et al., 1999; Ewers et al., 2006); however, immunogenicity was not reported. In our study also, we found PtfA to be present in all the five isolates signifying its role in pathogenesis of P. multocida.

Oma87 is a part of the outer membrane protein assembly complex, involved in assembly and insertion of beta-barrel proteins into the outer membrane. Omp87 (previously oma87) gene encoding Omp87 (87 kDa protein) was identified on the surface of all 16 serotypes of P. multocida isolates and considered as one of the immunodominant antigen (Kumar et al., 2013). Oma87 was also found to be present in all the five isolates of present study.

VacJ is a virulence-associated chromosome locus J and a surface lipoprotein of Shigella flexneri and Hemophilus influenzae. VacJ is a highly conserved and widely distributed outer membrane lipoprotein of P. multocida strains, which are known to affect a wide range of domestic as well as wild animals and birds. Structurally, majority of triacylated bacterial lipoproteins are considered to be similar in each bacterium and known to play important diverse roles in bacterial physiology and virulence (Shivachandra et al., 2014). VacJ gene was found to be present in our P52VAC, poultry and buffalo isolates.
4.3.2.1.5 Superoxide dismutase

SodA and SodC are the metalloenzyme superoxide dismutase (SOD), catalyses the conversion of superoxide radical anion to oxygen and hydrogen peroxide in the first of a series of protective reactions that remove cytotoxic free radicals generated during the reduction of molecular oxygen. Mechanisms to remove free superoxide are correspondingly widely found and three varieties of SOD being recognised. In the bacterial cytosol, iron- or manganese-cofactored SOD protects cellular proteins and DNA from free-radical attack. However, as the cytoplasmic membrane is impermeable to superoxide, these SODs cannot dismutate superoxide produced outside the cell, leaving the cell surface and periplasmic contents vulnerable (Lainson et al., 1996). Reactive oxygen and nitrogen species are the main actors of the innate defense responses of phagocytic cells. Accordingly, during macrophage infection, Salmonella enterica displayed an oxidative stress response illustrated by an increased expression of sodB, sodC and hmpA, the latter being implicated in the in vivo detoxification of nitric oxide (La et al., 2008). In our study, SodA and SodC genes were found in all the five isolates studied.

4.3.2.1.6 Sialic acid metabolism

Two genes, NanB and NanH were found in all the isolates studied using whole genome sequencing. Neuraminidase (sialidase) is prevalent among P. multocida strains and is expressed in vivo by P. multocida A:3 strains associated with bovine pneumonia (White et al., 1995; Ewers et al., 2006); both cell-bound and extracellular sialidases were reported (White et al., 1995). Its role in bovine respiratory disease remains largely unknown. Evidence of neuraminidase function in P. multocida infection suggested that the protein is protective in mice (Ifeanyi and Bailie, 1992), is associated with virulence (Muller and Krasemann, 1974) and may contribute to
bacterial adherence, colonization and persistence (Mizan et al., 2000). However, the relationship between the ability of *P. multocida* to produce the enzyme and to cause pneumonia in cattle has not been established. Two sialidase genes (*NanB* and *NanH*) were reported in 100 and 88.5% of *P. multocida* bovine strains investigated, respectively (Mizan et al., 2000). Mutation in *NanH* resulted in a bacterium that was not deficient in sialidase production but had reduced enzyme activity (Mizan et al., 2000).

Sialidases contribute to the virulence of pathogenic organisms which inhabit and invade mucosal surfaces (Schauer, 2009). The sialidases are enzymes that remove sialic acid conjugated to glycoproteins and glycolipids of eukaryotic cells (Hatfaludi et al., 2010). Sialic acid is then used as a carbon source by the bacteria, or it is incorporated into the outer membrane, aiding in the infectious process and in the inhibition of the host immune system (Vimr and Lichtensteiger, 2002). These enzymes also enhance bacterial virulence by unmasking key host receptors and reducing the effectiveness of host defences. Sialidases *NanH* and *NanB*, have been cloned and characterized from a fowl cholera isolate of *P. multocida* (Mizan et al., 2000). These sialidases differed in their specificity, with both able to utilize 2,3' sialyl lactose, but only *NanB* able to fully utilize 2,6' sialyl lactose. It was proposed that the presence of two sialidases with slightly different specificities would enhance the metabolic capacity of *P. multocida* in the host (Mizan et al., 2000). The uptake of sialic acid was also shown to be essential for virulence in mice (Steenbergen et al., 2005).

**4.3.2.1.7 Transcription regulation**

Fis is a transcriptional regulator, which has not previously been identified in virulence of *P. multocida*, was found to be present in all the isolates of *P. multocida*
stressed. Fis can act as both, a positive or negative regulator of transcription and it has both direct and indirect effects on gene transcription. The absence of functional Fis protein led to the loss of capsule expression in *P. multocida*. The presence of functional Fis protein is also required for the expression of virulence gene *pfhb2* and its predicted secretion partner *lspB2* in *P. multocida* (Steen et al., 2010).

Apart from these virulence associated genes, there were some other genes present in one or more isolates of *P. multocida*, which are directly or indirectly involved in disease progression and yet to be fully characterized in *P. multocida*, which are listed in Table 4.9.

**Table 4.9: List of probable virulence associated genes with their function**

<table>
<thead>
<tr>
<th>Genes</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>GlpE</em></td>
<td>Thiosulfate sulfurtransferase</td>
</tr>
<tr>
<td><em>GlpT</em></td>
<td>Glycerol-3-phosphate transmembrane transporter</td>
</tr>
<tr>
<td><em>CcmB</em></td>
<td>Heme exporter protein</td>
</tr>
<tr>
<td>Clp like protease</td>
<td>Degradation of misfolded proteins</td>
</tr>
<tr>
<td>CRISPR associated helicase Cas3 family</td>
<td>Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) is a prokaryotic defense mechanism against foreign genetic elements</td>
</tr>
<tr>
<td><em>DsbA, DsbC, DsbD, DsbE</em></td>
<td>Disulfide bond froming protein</td>
</tr>
<tr>
<td><em>NfuA</em></td>
<td>Fe/S biogenesis protein</td>
</tr>
<tr>
<td>Ferric transporter ATP binding subunit</td>
<td>Help in iron transport</td>
</tr>
<tr>
<td>Ferric uptake regulation protein</td>
<td>Help in iron transport</td>
</tr>
</tbody>
</table>
Results and Discussion...

<table>
<thead>
<tr>
<th>Gene/Protein</th>
<th>Function/Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>hemH</em></td>
<td>Ferrochelatase (Heme biosynthesis)</td>
</tr>
<tr>
<td>HSP 90, HSP 33, GrpE, hslo</td>
<td>Heat shock protein</td>
</tr>
<tr>
<td>Heme exporter protein B and D</td>
<td>Iron metabolism</td>
</tr>
<tr>
<td>HemN family oxidoreductase</td>
<td>Iron metabolism</td>
</tr>
<tr>
<td>YfcA</td>
<td>Inner membrane protein</td>
</tr>
<tr>
<td><em>fecB, fecC, fecD, fecE</em></td>
<td>Iron dicitrate transporter</td>
</tr>
<tr>
<td><em>MsmB</em></td>
<td>Cold shock like protein CspC</td>
</tr>
<tr>
<td><em>SanA</em></td>
<td>Encodes a dehydrogenase which is essential for Nikkomycin biosynthesis</td>
</tr>
<tr>
<td><em>SapD, SapF</em></td>
<td>Heme-iron utilization</td>
</tr>
</tbody>
</table>

*P. multocida* LPS is highly antigenic and capable of producing widespread vascular alterations and death, but it is poorly immunogenic (Heddleston and Rebers, 1975; Tsuji and Matsumoto, 1988; Ryu and Kim, 2000). Generally, *P. multocida* LPS is immunogenic or protective only when complexed with proteins (Ganfield *et al.*, 1976; Tsuji and Matsumoto, 1988; Ryu and Kim, 2000) or ribosomes (Phillips and Rimler, 1984). Ryu and Kim (2000) demonstrated complete protection in mice vaccinated with *P. multocida* type A:3 LPS-protein complex; however, neither the LPS or protein alone nor the LPS and protein as a mixture provided protection. The LPS-protein complex induced both humoral and cell-mediated immunity (Ryu and Kim, 2000). However, studies by Lu *et al.* (1991) failed to show protection in mice against *P. multocida* infection following passive immunization with affinity-purified...
rabbit anti-LPS serum. An opsonic but not bactericidal anti-LPS monoclonal antibody only partially protected against *P. multocida* infection in mice (Ramdani and Adler, 1991; Confer, 1993). LPS is also known as a stimulator of host defense against *P. multocida* infection (Iovane et al., 1998; Galdiero et al., 2000). LPS from a *P. multocida* serotype A strain enhanced humoral and cell-mediated immune responses in chickens (Maslog et al., 1999).

The gene sequences which are found to be present in the present study are the sequences which were being sequenced at a time of sequencing experiment and thus were definitely present. For absence of particular gene sequence in any of the isolates, the sequences were either absent or might not be sequenced at the time of experiment. The whole genome sequencing is based on random shearing and sheared DNA (~400 -500 bp) were used for the sequencing experiments. Since our sequences were in a scaffold and not in a single sequence, further study needs to be carried out for filling up the gaps using paired end sequencing approach. For the cattle isolate, further sequencing should be carried out for increasing in coverage and ultimately reduces the contigs for better prediction of genes/proteins. Any predictions should be confirmed by experimental data to definitively assign a definite biological function.

*P. multocida* strains of bovine, avian, and porcine origin have recently been sequenced and comparatively reviewed by Boyce *et al.* (2012). The authors noted, based on the nine genomes sequenced to date, that there was "no clear correlation between phylogenetic relatedness and host predilection or disease". Information is sparse on the location and characterization of the genes responsible for differences in virulence of avian and other *P. multocida*. None of the isolates was having Pasteurella multocida toxin (PMT) in their genome, which is specific to pig isolate of serotype D as well as recently sequenced buffalo isolate by Yap *et al.* (2013).
4.4 VACCINE DEVELOPMENT FOR P. multocida - A way forward

Live-attenuated micro-organisms, inactivated bacteria, purified microbial components, polysaccharide carrier protein conjugates, recombinant proteins or DNA are used as vaccines (Freddy et al., 2004). One of the most widely used vaccines in Asia is the whole cell formalin killed P. multocida P52 bacterin emulsified in aluminium hydroxide gel. This vaccine, however, does not provide long immunity (Qureshi and Saxena, 2014). It is therefore obvious that there are difficulties being encountered in producing vaccines for P. multocida, which may be due to the high antigenic variability and low reported work on whole genome sequence. In addition, the development of vaccines involves the detailed understanding of the mechanisms and mode of action of those virulent factors. Various studies have shown inconsistent efficacy of these vaccines (Moiser et al., 1994). Other challenges being encountered in the determination of a strong candidate gene from the pool of potential genes discovered, is the need to test the protective properties on animal model. Furthermore, potential gene products or protein will have to be purified and characterized before its antigenic properties can be fully known.

Recombinant antigens can be obtained from different expression systems after the targeted antigen has been determined. Recombinant proteins used as subunit vaccines have been reported safe as subunit vaccine (Hussaini et al., 2012). The cross-protective antigens still remain elusive (Shivachandra et al., 2011). Several candidate virulence genes were identified by Ahmad (2014) which includes \( \text{plpB, plpE, plp-40, PMT, OmpA, OmpH, Momp and PtfA} \). Other OMPs have been identified including \( \text{Omp16, Omp87 and iron-regulated OMPs such as TbpA, HemR, HbpA, HasR and HgbA} \) (Wheeler, 2009). Analysis of \( \text{OmpH} \) from different serotypes showed a high degree of conservation (Meredith et al., 2000). \( \text{OmpH} \) was shown to be of importance
in stimulating immunity to infection in animals (Davies and Lee, 2004). OmpA proteins are under evaluation as potential vaccine candidates (Shivachandra et al., 2013). The OmpA, will be a favorable gene for the development of naturally and recombinant-derived vaccines (Shivachandra et al., 2013).

The ptfA gene has also been proven to have a positive association with the disease outcome in cattle and hence could be used to develop broad cross-protective and disease specific subunit vaccine along with suitable adjuvants especially for HS in bovines and septicaemic pasteurellosis in sheep and goat (Shivachandra et al., 2013). Controversies have surrounded the acceptance of the plpB as a candidate gene for the development of P. multocida vaccine. It was first purified and proved to confer protection against different serotypes of P. multocida (Rimler, 2001). The protein was confirmed to be P. multocida lipoprotein B by Tabatabai and Zehr (2004). Other reported candidate genes include the AroA gene, as deletion of AroA (A gene encoding the P. multocida toxin) of serotype B2 was used as an effective vaccine against HS (Ataei et al., 2009).

Virulent genes tbpA and pfhA as well as capsule biosynthesis genes are supposed to be important epidemiological marker genes for characterizing P. multocida field strains (Ataei et al., 2009). Study by Ewers et al. (2006) confirmed that tbpA and pfhA are virulent genes. TbpA gene (82 kDa protein) first isolated in HS strains of capsule serotype B: 2, 5 in buffaloes and cattle (Veken et al., 1994) have been shown to have strong homology with A1 and D1 (Shivachandra et al., 2005). Verma et al. (2013) found high occurrence of pfhA and tbpA genes among P. multocida isolates from diseased as well as healthy cattle. Some virulence associated genes such as sodA and sodC were found in a higher percentage among isolates from diseased animals as compared to isolates obtained from apparently healthy animals.
Shivachandra et al. (2013) suggested that a recombinant OmpH might be a useful vaccine candidate. Surface exposure and high number of copy makes the OmpA-like proteins favourite candidates for vaccine development (Shivachandra et al., 2013). Synthetic porin (OmpH) peptides have been shown to induce limited protection against P. multocida (Luo et al., 1999).

Several genes involved in virulence and immunity in P. multocida were localized and targeted as having potential in the development of subunit vaccines (Fuller et al., 2000; May et al., 2001; Hatfaludi et al., 2010). Type 4 fimbriae (PtfA), capsule, OMP87 (Omp87), fragments of exotoxin (PMT), porin (OmpH), lipoprotein E (plpE), heme-binding proteins (HasR, HemR and HgbA), filamentous hemagglutinin (fhaB2), adhesin (OmpA) and transferrin binding protein (TbpA) have been considered as candidate antigens (Veken et al., 1994; Ruffolo et al., 1997; Paustian et al., 2001; Bosch et al., 2004; Prado et al., 2005; Shivachandra et al., 2005; Wu et al., 2007; Hatfaludi et al., 2010). Recombinant proteins used as subunit vaccines were safe for immunocompromised animals as they did not cause disease (Verma and Jaiswal, 1998; De Alwis, 1999).

The observations that bovine strains of P. multocida regularly harbored several virulence genes (Type 4 fimbrial gene ptfA, iron acquisition genes tbpA, hgbA, hgbB, HasR, and genes for multiple hemoglobin-binding proteins), independent of its capsule type and these genes are highly expressed in the host, which strongly suggest that their protein products are potential subunit vaccine candidates (Veken et al., 1996; Ruffalo et al., 1997; Bosch et al., 2004). A surface exposed, highly immunogenic lipoprotein (plpE) of P. multocida conferred 80-100% protective immunity in mice and induced cross-serotype protective immunity (Wu et al., 2007).
At present, there is no commercial recombinant HS subunit vaccine and cross-protective antigens still remain elusive.

This study is apparently the first attempt in India involving local *P. multocida* isolates from four different species and a vaccine strain for the purpose of identifying virulence/virulence associated genes using modern biotechnological tools like pyrosequencing based whole genome sequencing. The study aids in data of whole genome sequencing of bacterial pathogens particularly for *P. multocida* and also provides new insight into their genomic characters and possible molecular mechanisms involved in disease process. The present findings would provide a much needed base for further screening of virulence associated genes and identification of certain markers for early diagnosis as well as characterization of *P. multocida*, which continues to pose challenges as a menace against the health management of animals. Genes which have been found in all the isolates under the study can be explored as specific probes for the early diagnosis of the *P. multocida* disease. Further, future scientific endeavors targeting the vaccine design for *P. multocida* may get a scientific support from this data, so as to formulate modern and more effective vaccines, for better animal health.
SUMMARY AND CONCLUSIONS
CHAPTER V

SUMMARY AND CONCLUSIONS

5.1 SUMMARY

Haemorrhagic septicemia is an important disease of cattle, buffaloes, poultry and other livestock species. For effective control of the disease, an efficacious and longer duration immunity vaccine is required. Identification and characterization of important immunogens of the bacteria help in designing an improved vaccine as well as identification of the OMPs and exploiting it would serve as a candidate for development of diagnostics.

Genome sequencing has revolutionized approach for understanding bacterial pathogenesis. As for obligate intracellular bacteria and other microbial species that cannot be cultured in vitro, genomics has provided virtually the only path for efficient discovery of virulence genes. Genomics study has greatly benefited research on organisms that can be cultured by increasing the pace of identification and characterization of virulence factors. Improved understanding of disease etiology helps to direct research into therapies. Genomics is a promising tool for investigating the differences between invasive and non-invasive bacteria at the population as well as at host levels.

Next generation sequencing technologies empower the use of highly parallel methodologies that allow study of all the genes or all the proteins of a pathogen in the context of its host or under various physiological or genetic states of interest. *P. multocida* whole genome sequences and advances in ability to genetically manipulate the organism will facilitate major advances in understanding of disease pathogenesis. Determining the *P. multocida* genes involved in bacterial adherence, as well as other
virulence associated genes, can assist in understanding of the molecular pathogenesis of *P. multocida*. Changes in host-pathogen interaction resulting from changes in environmental signals and the associated changing molecular mechanisms for virulence are most likely the causes of differences between commensal and pathogenic *P. multocida*.

In spite of its undesirably high cost at present, sequence-based identification of the virulent isolates is a convenient alternative and is recommended when unusual or atypical isolates are observed. It can also be used in molecular epidemiological study and for selecting/designing a suitable vaccine candidate.

For Whole genome sequencing of *P. multocida*, different isolates belonging to different species viz. P52 (vaccine strain), poultry (Anand1_poultry), goat (Anand1_goat), buffalo (Anand1_buffalo) and cattle (Anand1_cattle) were collected. All the isolates were confirmed based on cultural and biochemical characteristic as well as PM-PCR based identification method and found to be of *P. multocida*. For whole genome sequencing of all the five isolates, DNA was extracted using phenol-chloroform method. Extracted DNA were quality and quantity checked for confirming purity and nebulized using nitrogen gas. Sheared DNA fragments ends were polished using T4 DNA polymerase and Taq DNA polymerase. MIDs adaptors were ligated to each of the isolates and small fragments were removed using AMPure beads. dsDNA library prepared were quantity checked using TBS fluorometer as well as with High Sensitivity DNA chip. Small volume Emulsion PCR was done to check which DNA molecules/beads ratio gives recovery of 6-12% after enrichment. Large volume PCR was done using 4.0 molecules/beads recovery (i.e. 9%). After recovering positively clonally amplified DNA beads, sequencing primers were attached to each of the fragments and sequined on 454 GS FLX Titanium whole genome sequencer.
Sequencing of DNA resulted into 118843, 113997, 105729, 134886 and 31346 reads with 42,598,100 (42.59Mb), 29,000,497 (29.00Mb), 21,890,353 (21.89Mb), 39,756,349 (39.75Mb) and 7,429,658 (7.42Mb) sequence bases for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. The coverage obtained were 18.87, 12.85, 9.70, 17.61 and 3.29 for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively.

Reference mapping of reads with the first sequenced genome of *P. multocida* i.e. of Pm70, an avirulent strain isolated from poultry in Australia, was done using GS Reference Mapper and generated 38,079,806 (89.52%), 20,085,356 (87.38%), 19,867,143 (90.81%), 25,095,466 (63.22%) and 6,145,156 (82.87%) mapped bases with 105327, 97674, 95092, 86765 and 24967 mapped reads for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. Mapping results generated 155, 402, 257, 227 and 1998 contigs for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. Remaining reads which were not mapped to Pm70 were *De novo* assembled using *De novo* Assembly software to search for sequences associated with plasmid. None of the *De novo* assembled sequences matched with the plasmid using BLASTX.

For genome annotation, two different pipelines were used i.e. Rapid Annotation using Subsystem Technology (RAST) and Prokaryotic Genome Automatic Annotation Pipeline (PGAAP). RAST defined 27 major subsystems, out of which, two subsystems for ‘photosynthesis’ as well as ‘motility and chemotaxis’ were not assigned to any of the five isolates as expected. RAST analysis revealed genome size of 2,273,366bp (2.27Mb), 2,227,943bp (2.22Mb), 2,285,382bp (2.28Mb), 2,045,610bp (2.04Mb) and 1,438,517bp (1.43Mb) with 68, 54, 54, 40 and 0 RNA for
P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively.

RAST analysis revealed maximum subsystems assigned to ‘amino acids and derivatives’, ‘carbohydrates’, ‘protein metabolism’ and ‘cofactor and vitamins, prosthetic groups and pigments’. Cattle isolate was omitted from the RAST comparison as very less subsystems were assigned with lower counts because of very less coverage (3.29X). Under ‘virulence, disease and defence’ subsystem, presence of DedA, DedD and toxin under colicin and bacteriocin production was noted in P52 vaccine strain, poultry and goat isolates. Genes gyrA, gyrB, Parc and ParD under ‘resistance to fluoroquinolones’ subsystem were present in all the other four isolates. There was also presence of negative regulator of betalactamase expression, BLR gene leading to resistance expressed by this organism as well as multidrug resistance efflux pump cluster genes, MATE (Multidrug and toxin extrusion), MacA and MacB (Macrolide specific efflux protein) in P52 vaccine strain, poultry and goat isolates.

Second annotation was done using PGAAP pipeline developed by NCBI using reads more than 200bp in length. PGAAP analysis revealed genome size of 2,273,366bp (2.27Mb), 2,227,943bp (2.22Mb), 2,285,382bp (2.28Mb), 2,045,610bp (2.04Mb) and 1,438,517bp (1.43Mb) with 40.40%, 40.20%, 40.50%, 40.90% and 41.00% of G+C contents for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively. Total number of coding sequences (CDS) were 2066, 2337, 2319, 3258 and 3623 and total number of protein encoding genes (PEG) were 2194, 2284, 2266, 3218 and 3590 respectively, for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle. Total no. of RNA assigned were 64, 53, 53, 41 and 33 for P52VAC, Anand1_poultry, Anand1_goat, Anand1_buffalo and Anand1_cattle, respectively.
Based on PGAAP analysis, 55 genes were identified and grouped under seven categories viz. capsule, fimbriae and adhesion, iron metabolism, outer membrane protein, superoxide dismutase, sialic acid metabolism and transcription regulation. Genes SodA and SodC under superoxide dismutase, NanH and NanB under sialic acid metabolism and Fis under transcription regulation category were present in all the five isolates of *P. multocida*. Remaining 50 genes were present in varying frequency either in one or more isolates.

For capsule production, there were nine genes identified, out of which, genes PglA and Kmt1 were present in all the five isolates. Gene HyaE was present only in goat isolate, whereas LctP gene was present in goat and cattle isolates only. Genes HexA and HexC were absent in cattle and buffalo isolates, while HexB and HexD were absent in goat as well as cattle isolates.

Under fimbriae and adhesion category, there were 16 genes, of which Hsf, PfhB2, PfhR, PfB, PlpB, and Plp4 genes were present in all the five isolates. In P52 vaccine strain, HofC and ComE genes were absent; TadE, RcpA and RcpB were absent only in buffalo isolate. Gene PlpE and TadF were absent in buffalo as well as in cattle isolate. Genes PlpP and PfhB1 were absent in cattle isolate whereas, ClpB was absent in P52 as well as in cattle isolate of *P. multocida*.

In iron metabolism category, 16 genes were present, out of which, ExbB, FbpB, HbpA, HgbA, HemU, OmpW, Rfb and RffG genes were present in all the five isolates of *P. multocida*. For cattle isolate, FbpA and TonB genes were absent, where as in goat isolate, TbpA and TonB dependent lactoferrin and transferrin receptor were absent. Gene FbpC was absent in buffalo isolate. TonB-dependent receptor was
Summary and Conclusions...

present in poultry and goat isolates only. Translocation protein TolB was present and
\textit{HemR} absent only in P52 vaccine strain.

For the category of outer membrane proteins, \textit{HasR}, \textit{LppB}, \textit{LspB}, \textit{OmpH} and
\textit{PtfA} genes were present in all the five isolates of \textit{P. multocida}, out of 9 genes falling
in this category. \textit{VacJ} gene was absent in goat and cattle isolates, whereas \textit{Oma87}
gene was absent only in poultry isolate. Gene \textit{PfhA} was present only in P52 vaccine
strain. Outer membrane protein \textit{LoIB}, was found present in P52 vaccine strain as well
as in cattle isolate.

Based on the results obtained in this study, following conclusions could be
drawn:

5.2 CONCLUSIONS

5.2.1 The genome size of \textit{Pasteurella multocida} vaccine strain (P52), poultry,
goat, buffalo and cattle isolates was 2.27Mb, 2.22Mb, 2.28Mb, 2.04Mb
and 1.43Mb. The genomes of vaccine strain and the goat isolate were little
larger than the reference Pm70 (2.25Mb) genome, while the genome of the
poultry and buffalo isolates were smaller.

5.2.2 GC contents of the \textit{P. multocida} vaccine strain (P52), poultry, goat,
buffalo and cattle isolates were 40.40%, 40.20%, 40.50%, 40.90% and
41.00%, respectively and number of RNA genes were 64, 53, 53, 41 and
33.

5.2.3 The protein encoding genes (PEG) were 2194, 2284, 2266, 3218 and 3590
and the number of coding sequences (CDS) were 2066, 2337, 2319, 3258
and 3623 for \textit{P. multocida} P52, poultry, goat, buffalo and cattle isolates,
respectively.
5.2.4 Based on RAST analysis, maximum number of subsystems were assigned to 'amino acids and derivatives', 'carbohydrates', 'protein metabolism' and 'cofactor and vitamins, prosthetic groups and pigments'. No subsystems was assigned to 'photosynthesis' and 'motility and chemotaxis' group.

5.2.5 By PGAAP analysis, SodA and SodC genes under superoxide dismutase category, NanH and NanB genes under sialic acid metabolism category and Fis gene under transcription regulation category were present in all the five isolates of P. multocida.

5.2.6 Gene HyaE, which codes for hyaluronic acid and involved in capsular assembly was found only in the goat isolate. Possibility of targeting this gene for detecting P. multocida of goat origin should be further explored.

5.2.7 Gene ComE under fimbrie and adhesion category and HemR under iron metabolism category were absent only in P52 vaccine strain of P. multocida, whereas, genes RcpA, RcpB and TadE under the same category were absent only in buffalo isolate.

5.2.8 Type IV fimbriae encoding Plp4 gene was uniquely present in all the five isolates and this might apparently be the first report of Plp4 gene from P. multocida of ruminants using whole genome sequence.

5.2.9 Under iron metabolism category, gene FbpC was absent only in buffalo isolate. Genes ThpA and TonB-dependent lactoferrin and transferrin were absent only in goat isolate. Protein TolB was present only in P52 vaccine strain.
5.2.10 For outer membrane proteins, gene \( PfhA \) and outer membrane protein LolB were present only in P52 vaccine strain, whereas \( Oma87 \) was present only in the poultry isolate.

5.2.11 Fis, a transcriptional regulator identified recently in the virulence of \( P.\ multocida \) was found to be present in all the five isolates (i.e. P52, poultry, goat, buffalo and cattle).

5.2.12 Gene \( VacJ \), which codes for outer membrane lipoprotein of \( P.\ multocida \) and playing role in bacterial physiology and virulence was found in P52 vaccine strain, poultry and buffalo isolates.

5.2.13 Genes which are found commonly present in all the isolates regardless of the serotype as well as the genes found exclusively present or absent in the isolate of a particular species can be further explored as specific probe for the detection as well as vaccine designing for \( P.\ multocida \).


Paczian, T.; Parrello, B.; Pusch, G. D.; Reich, C.; Stevens, R.; Vassieva, O.;

sequencers as a tool for genotyping of highly polymorphic multilocus MHC

Babu, M. M.; Priya, M. L.; Selvan, A. T.; Madera, M.; Gough, J.; Aravind, L. and
functional assignments to predicted lipoproteins. *J. Bacteriol.* 188(8): 2761-
2773.

*Pasteurella multocida* toxin activates human monocyte-derived and murine
bone marrow-derived dendritic cells *in vitro* but suppresses antibody

Haemorrhagic septicaemia. FAO Animal Production and Health Paper 33: 11-
33.

in Fis levels upon nutrient upshift in *Escherichia coli*. *J. Bacteriol.* 174(24):
8043-8056.


genes of *Pasteurella multocida* isolated from cases of fowl cholera by

transmigration through bovine endothelial cells *in vitro* by protein H and LPS
of *Pasteurella multocida*. Immunobiology 202(3): 226-238.

Berkowitz, A.; Perez, N.; Clubb, S.; DeRisi, J. L.; Ganem, D. and Lublin, A.
(2009). Experimental induction of proventricular dilatation disease in
cockatiels (*Nymphicus hollandicus*) inoculated with brain homogenates

Ganfield, D. J.; Rebers, P. A. and Heddeleston, K. L. (1976). Immunogenic and toxic
properties of a purified lipopolysaccharide-protein complex from *Pasteurella

Garcia, N.; Fernandez-Garayzabal, J. F.; Goyache, J. Dominguez, L. and Vela, A. I.
(2011). Associations between biovar and virulence factor genes in *Pasteurella

Garrido, M. E.; Bosch, M.; Medina, R.; Bigas, A.; Llagostera, M.; Perez De Rozas, A.
M.; Badiola, I. and Barbe, J. (2003). fur-independent regulation of the
*Pasteurella multocida hbpA* gene encoding a haemin-binding protein.
Microbiology 149(8): 2273-2281.


genetic changes enhance virulence of methicillin-resistant and sensitive Staphylococcus aureus. BMC Microbiology 7: 99


required for non-specific adherence of *Actinobacillus actinomycetemcomitans.*

*Mol. Microbiol.* **40:** 542-554.


References...


References


Morton, D. J.; Madore, L. L.; Smith, A.; Vanwagoner, T. M.; Seale, T. W.; Whitby, P. W. and Stull, T. L. (2005). The heme-binding lipoprotein (HbpA) of
References...


Roehrig, S. C.; Tran, H. Q.; Spehr, V.; Gunkel, N.; Selzer, P. M. and Ullrich, H. J. (2007). The response of *Mannheimia haemolytica* to iron limitation:


Typhimurium contribute to the virulence of the organism. *Infect. Immun.* 72(7): 3987-4003.


References...


References...


*** *** ***