Complete genomic sequence of *Pasteurella multocida*, Pm70

Barbara J. May*, Qing Zhang*, Ling Ling Li*, Michael L. Paustian*, Thomas S. Whittam†, and Vivek Kapur*‡

*Department of Veterinary Pathobiology, University of Minnesota, St. Paul, MN 55108; and National Food Safety and Toxicology Center, Michigan State University, East Lansing, MI 48824

Communicated by Harley W. Moon, Iowa State University, Ames, IA, December 29, 2000 (received for review November 14, 2000)

We present here the complete genome sequence of a common avian clone of *Pasteurella multocida*, Pm70. 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* suggest that *P. multocida* and *H. influenzae* diverged ≈270 million years ago and the γ subdivision of the proteobacteria radiated about 680 million years ago. Two previously 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. Overall, the complete genomic sequence and preliminary functional analyses provide a foundation for future research into the mechanisms of pathogenesis and host specificity of this important multispecies pathogen.

It has been more than a century since Louis Pasteur conducted experiments with *Pasteurella multocida* (*Pm*), first demonstrating that laboratory attenuated bacteria could be used for the development of vaccines (1). Despite this seminal discovery, the molecular mechanisms for infection and virulence of *Pm* have remained largely undetermined, and this organism has continued to cause a wide range of diseases in animals and humans. It is the causative agent of fowl cholera in domesticated and wild birds, hemorrhagic septicemia in cattle, atrophic rhinitis in swine, and is the most common source of infection in humans because of dog and cat bites (2, 3).

Because 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. To better understand the molecular basis for *Pm*'s virulence, pathogenicity, and host specificity, we sequenced the genome of a common avian isolate recovered from a recent case of fowl cholera in chickens. The analysis identified a total of 2,014 open reading frames, including several encoding putative virulence factors. In particular, *Pm* has two genes with significant homology to the filamentous hemagglutinin gene in *Bordetella pertussis*, as well as more than 50 genes with a potential role in iron uptake and metabolism. The analysis also provides strong evidence that *Pm* and its close relative, *Hemophilus influenzae* (*Hi*), diverged ≈270 million years ago (mya) and that the γ subdivision of the proteobacteria, a group that contains many of the pathogenic Gram-negative organisms, radiated ≈680 mya.

Materials and Methods

Multilocus Enzyme Electrophoresis (MLEE). MLEE was used to determine the population genetic structure of *Pm* from avian sources. Methods for MLEE have been described in detail elsewhere (4). Briefly, 271 *Pm* isolates recovered from 14 wild and domesticated avian species from throughout the world were grown and sonicated for the collection of the enzyme-containing supernatant. Histochemical staining for 13 metabolic enzymes [mannose phosphate isomerase, glutamate dehydrogenase, shikimic acid, glucose-6-phosphate dehydrogenase, nucleoside phosphorylase, phosphoenol pyruvate, malate dehydrogenase, fumarase (two isoforms), phosphoglucose isomerase, adenylate kinase, 6-phosphogluconate deyhdrogenase, and mannitol-1 phosphate deyhdrogenase] was conducted to determine distinct mobility variants of each enzyme. Each isolate was characterized by its combination of alleles at 13 enzyme loci, and electrophoretic type designations were assigned by computer analysis as described (4).

Sequencing and Assembly. A shotgun strategy (5) was adopted to sequence the genome of Pm70 (capsular type A; serotype 3). To create a small (1.8- to 3.0-kb) insert library, genomic DNA was initially isolated by using a chloroform/cetyltrimethylammonium bromide-based method, as described (6). Next, the DNA was sheared by nebulization (http://www.genome.ou.edu) and cloned into pUC18 for isolation and sequencing. Approximately 25,000 clones were sequenced from both ends by using Dye-terminator chemistry on ABI 377 (Applied Biosystems) and Megabace (Molecular Dynamics) sequencing machines. A total of 53,265 sequences were used to generate the final assembly, giving about a 10-fold coverage of the genome. Sequence assembly and verification were accomplished by using PHRED-PHRAP (P. Green, http://genome.washington.edu). To close the approximately 100 gaps at the end of the shotgun phase, several methods were used, including primer walking, homology-based comparisons of gap ends, and multiplex PCR. Primer walking was performed on genomic DNA or on a large-insert λ library (15–20 kb). The final assembly was compared to a previously published genetic map of *Pm* (7).

Potential coding sequences (CDSs) were predicted by using ORPHEUS, GLIMMER, and ARTEMIS, and the results were compared and verified manually in ARTEMIS (8, 9). Homology studies were completed with BLASTP analysis by using a database constructed by the Computational Biology Center at the University of Minnesota (http://www.cbc.umn.edu). Transfer RNAs were predicted by TRNASCAN-SE (10).

Abbreviations: mya, million years ago; CDS, coding sequence; *Pm*, *Pasteurella multocida*; *Hi*, *Haemophilus influenzae*; *Ec*, *Escherichia coli*; MLEE, multilocus enzyme electrophoresis; D, distance.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AE004439).

‡To whom reprint requests should be addressed at: Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Avenue, St. Paul, MN 55108. E-mail: vkapur@umn.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Microarray Analysis and Isolation of Total RNA. A flask of brain–heart infusion (BHI) media (Becton Dickinson) was inoculated with *Pm* Pm70 and grown to log phase (OD_{600} = 0.7) at 37°C. The culture was then split into two 180-ml volumes, which were pelleted at 4°C, washed with 1× PBS, and pelleted again. One pellet was resuspended in 180 ml BHI and the other in 180 ml BHI containing the iron-chelator 2,2′-dipyridyl (200 μM). The cultures were incubated on a rotary shaker at 37°C, and 30-ml volumes were removed at time points of 15, 30, 60, and 120 min after resuspension. The cultures were briefly centrifuged, and the pellets were flash frozen in dry ice and ethanol. Total RNA extractions were performed with RNeasy maxi columns (Qiagen, Chatsworth, CA), with DNase digestions done on-column by adding 82 Kunitz units of enzyme (Qiagen) and incubating at room temperature for 15 min. Ten micrograms of each RNA sample was used in two separate hybridization experiments on identical arrays.

cDNA Synthesis and Labeling Conditions. Construction of cDNA and DNA microarray hybridization approaches were performed as described at our web site (http://www1.umn.edu/agac/microarray/protocols.html). In brief, 30 μg of random hexamers (Amersham) and 10 μg of total RNA were initially preheated at 70°C for 10 min, quick-cooled on ice for 10 min, and incubated for 2 h at 42°C with the following reverse transcription–PCR reaction mix: first-strand buffer, 10 mM DTT; 380 units of Superscript II RT; and 500 μM dATP, dCTP, dGTP; 100 μM dTTP, all from Life Technologies–GIBCO/BRL; and 400 μM amino-allyl-labeled dUTP (Sigma; 4:1 ratio). After hydrolysis with 10 μl 1 M NaOH and 10 μl 0.5 M EDTA for 15 min at 65°C, the samples were neutralized with the addition of 25 μl of 1 M Tris·HCl (pH 7.4) and cleanup was performed with Microcon 30 filters (Millipore). The fluorescent monofunctional *N*-hydroxysuccinimide-ester dyes Cy3 and Cy5 were coupled with the cDNAs from control and iron-depleted bacteria, respectively for 1 hour, quenched with 4.5 μl hydroxylamine (4 M; Sigma), and cleaned with the Qia-Quick PCR Purification Kit (Qiagen). The samples were then dried down and stored for no longer than 24 h at 4°C before use.

DNA Microarray Hybridization and Analysis. Preparation of DNA microarrays (including primer sequences and PCR protocols), and hybridization techniques are also described at our web site mentioned above. Briefly, a library of targets representing all 2,014 open reading frames (ORFs) from *Pm* Pm70 was constructed with primers designed to amplify fragments ≤500 base pairs from each ORF from genomic DNA template. Two rounds of PCR were performed to minimize genomic DNA contamination, and the final 100-μl reactions were checked for quality on agarose gels and purified with MultiScreen PCR plates (Millipore). Arrays of 1,936 ORF segments representing the successful amplifications and >95% of the *Pm* CDSs were printed by using a Total Array System robot (BioRobotics, Cambridge, UK). Samples containing no DNA were also spotted as a negative controls.

Hybridization was conducted by resuspending our cDNA-labeled probes in dH_2O, 3× SSC, along with SDS and salmon sperm to be used as blockers. The probe was then added to the array and incubated at 63°C for 6 hours. Images of the hybridized arrays were obtained with a ScanArray 5000 microarray scanner, and fluorescent intensities were analyzed by using the program QuantArray (GSI Lumonics, Watertown, MA). Hybridization experiments were repeated two times to ensure reproducibility of the results. To obtain a Cy3/Cy5 intensity ratio for each spot, the intensities for each spot were first standardized by subtracting local background and normalizing individual spot intensities against the total spot intensity for both channels. Spots with a 2-fold or greater intensity ratio were further analyzed for their differences in gene expression. Hierarchical clustering and analysis were performed by using the publicly available programs CLUSTER and TREEVIEW (http://www.microarrays.org/software; ref. 11).

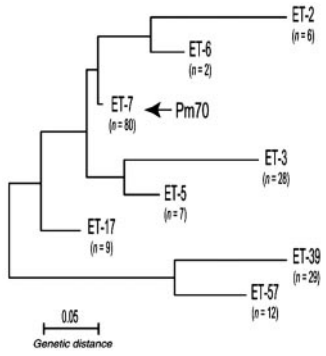

Fig. 1. Genetic relationships among the eight common clones of *Pm* recovered from avian sources. A total of 271 independent isolates of *Pm* recovered from 14 avian species were characterized for differences in electrophoretic mobility at 13 loci by MLEE. Of the 82 electrophoretic types that were identified, more than two-thirds of the isolates were represented by the eight clones shown in the dendrogram. Furthermore, a total of 31% of the isolates were represented by a single group, electrophoretic type 07 (arrow) that included the isolate Pm70.

MICROBIOLOGY

Results and Discussion

***Pm* Population Genetic Structure.** To identify the bacterial clones responsible for most infections in avian hosts, we first determined the genetic diversity and population structure of *Pm* strains collected from a wide range of wild and domesticated birds. A total of 271 independent isolates recovered from 14 avian species were characterized for protein polymorphisms at 13 enzyme loci by multilocus enzyme electrophoresis (4). The analysis revealed 72 electrophoretic types (ETs), with an average genetic diversity (H) of 0.474 ± 0.049, indicating that considerable genetic variation exists within *Pm*. However, more than two-thirds (69%) of all isolates were accounted for by only eight clones (ETs), and 31% of the isolates were of a single clone designated ET-7 (Fig. 1). Strain Pm70, isolated in 1995 from a case of fowl cholera in chickens, is a member of this *Pm* clone and was thus targeted for genome sequence analysis.

Characteristics of the Pm70 Genome. We used the random shotgun approach to generate more than 53,000 sequence fragments from strain Pm70, which were then assembled into a single circular sequence of 2,257,487 base pairs (Fig. 2). The putative origin of replication of the chromosome was identified on the basis of the presence of dnaA boxes, characteristic oligomer skew, and G-C skew immediately before the first putative coding sequence, *gidA* (12, 13). The entire sequence specifies 2,014 potential CDSs with an average size of 998 base pairs, which in sum account for 89% of the entire chromosome (Fig. 7, which is published as supplemental data on the PNAS web site, www.pnas.org). The remaining 11% of the sequence consists of 6 complete rRNA operons (16S-23S-5S), 57 tRNA genes representing all 20 amino acids, and a relatively small number of noncoding elements (Table 1). Sequence comparisons identified 200 CDSs (10%) unique to *Pm* and 1,197 CDSs with orthologs in both the *Hi* and *Escherichia coli* (*Ec*) genomes. These results are consistent with an overall close relationship and support the classification of these bacteria in the γ subdivision of the proteobacteria. Because 26% of the CDSs are most similar to hypothetical proteins of unknown functions, as is seen in other

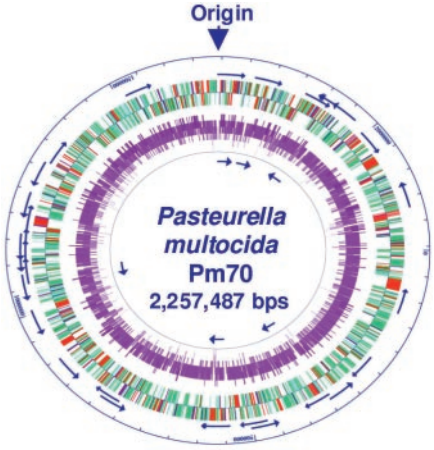

Fig. 2. Circular representation of the genome of *Pm*, Pm70. The *Pm* genome and its coding regions with homologies, the tRNA and rRNA operons, and the overall G-C content are presented. The outer circle represents the scale in base pairs with the origin noted. The outer arrows represent the 57 tRNAs and the inner arrows represent the 6 complete rRNA operons (16S–23S–5S). The 2,014 potential coding sequences are represented by colors depicting homology to both *Ec* and *Hi* (green), *Ec* alone (purple), *Hi* alone (light blue), or to organisms other than *Hi* or *Ec* or unique to *Pm* (red). The G-C % of each coding sequence is represented in the interior circle by different shades of pink in increments of 5%: the lightest pink represents a G-C % of 25–30% whereas the darkest pink represents 45–50%. The figure was generated by using GENESCENE software (DNAstar, Madison, WI).

completed microbial genomes, a substantial portion of *Pm*'s biochemistry and cell biology remains to be discovered (14).

The *Pm* genome encodes complete sets of enzymes for the oxidative pentose phosphate and Entner–Doudoroff pathways, glycolysis, gluconeogenesis, and the trichloroacetic acid cycle [trichloroacetic acid (TCA) cycle; see Table 3, which is published as supplemental data on the PNAS web site]. These pathways are also present in both *Ec* (15) and *Hi* (5), except for the absence of a complete TCA cycle in *Hi*. The *Pm* genome, like *Hi* and *Ec*, encodes a complete set of instructions for the synthesis of all 20 amino acids and purine and pyrimidine nucleotides. In contrast to *Ec*, catabolic pathways for several amino acids, including those for asparagine, histidine, leucine, lysine, and phenylalanine, are missing in both *Pm* and *Hi*. Pathways involved in sulfur uptake and metabolism as well as nitrogen and folic acid metabolism are present in the bacterium. Strains of *Pm* are known to use a number of carbohydrates, including sucrose, fructose, galactose, mannose, sorbitol, and ethanol, which is consistent with the presence of corresponding genes identified in our analysis (16). On the basis of the presence of identifiable orthologs, several additional carbohydrates, such as maltotriose, fucose, ribose, and sorbose, are also likely to be used by *Pm*.

Table 1. Summary of the complete genome of *P. multocida*

	No. of CDS	%
Number of CDS	2,014	
Homologues to peptides	1,814	90
Homologues to hypothetical proteins	531	26
CDS unique to *Pm*	200	10
Homologues to *Hi*	1,421	71
Homologues to *Ec*	1,392	69
Orthologues to *Ec* but not *Hi*	223	11
Orthologues to *Hi* but not *Ec*	195	10
CDS genome coverage		89
Overall GC%		41
Average molecular weight of proteins	37,081 K_d	
Average gene size	997 bp	
tRNAs	57	
rRNA operons	6	

Table 2. Evolutionary distances between *P. multocida, H. influenzae,* and *E. coli*

	D* values for CDSs	
Comparison	Conserved (n = 774)	Divergent (n = 243)
Pm-Hi	24.0 ± 11.9	68.4 ± 13.8
Pm-Ec	59.1 ± 31.0	162.4 ± 79.9
Hi-Ec	61.3 ± 31.8	170.7 ± 80.6

*D, evolutionary distance for each pairwise comparison.

Evolutionary Relationships of *Pm*. Members of the *Pasteurella*, *Haemophilus*, and *Escherichia* genera belong to the γ subdivision of the proteobacteria and include many of the most pathogenic Gram-negative bacteria. The availability of the complete genome sequence from three members of this subdivision provided us with the ability to examine the comparative evolution of these organisms at a genomic scale. To determine evolutionary relationships between *Pm*, *Hi*, and *Ec*, the 1,197 CDSs represented in all three organisms were aligned by CLUSTALW, and the percentage of identical amino acids in aligned sequences was determined. We next calculated evolutionary distance (D) per Grishin (17) and tabulated the distribution of D for each pairwise comparison of *Pm, Hi*, and *Ec* for the 1,197 orthologs. The mean (±standard deviation) of the pairwise distances were: *Pm-Hi* 56.6 ± 83.6; *Pm-Ec* 104.0 ± 89.1; and *Hi-Ec* 114.5 ± 103.6. All three distributions were J-shaped with long tails, and separation of the orthologs outside the 1.5 interquartile range resulted in the identification of 243 divergent and 774 conserved CDSs (Table 2). Scatter plots of pairwise D values show near linear relationships (Fig. 3), indicating remarkable consistency in the rate of divergence of orthologous genes among these genera. Feng and Doolittle (18) calibrated Grishin's D against the fossil

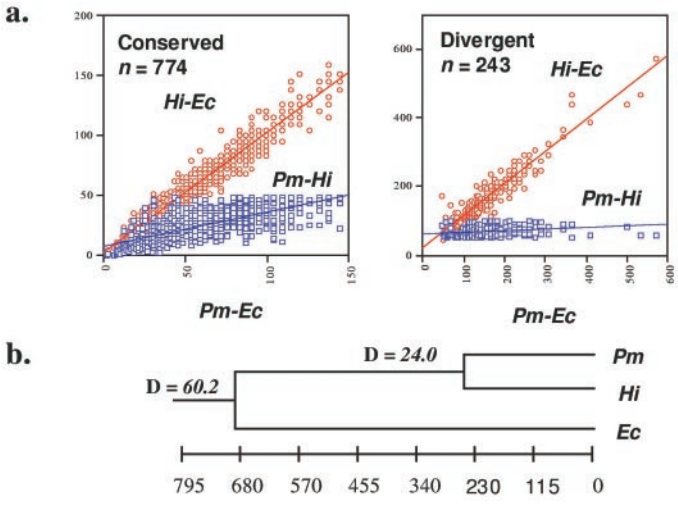

Fig. 3. Genome-scale evolutionary comparisons between *Pm*, *Hi*, and *Ec*. Scatter plots were created from pairwise evolutionary distance (D) values, as described. The *x* axis represents the pairwise D value for *Pm* and *Ec* and the *y* axis represents the pairwise D value for either *Hi* and *Ec* or *Pm* and *Hi* as indicated. (*a*) Scatterplot for the 774 conserved sequences and 243 divergent sequences. (*b*) Dendrogram showing the evolutionary distance in millions of years between *Pm*, *Hi*, and *Ec*. D values for each of the branch points are also presented.

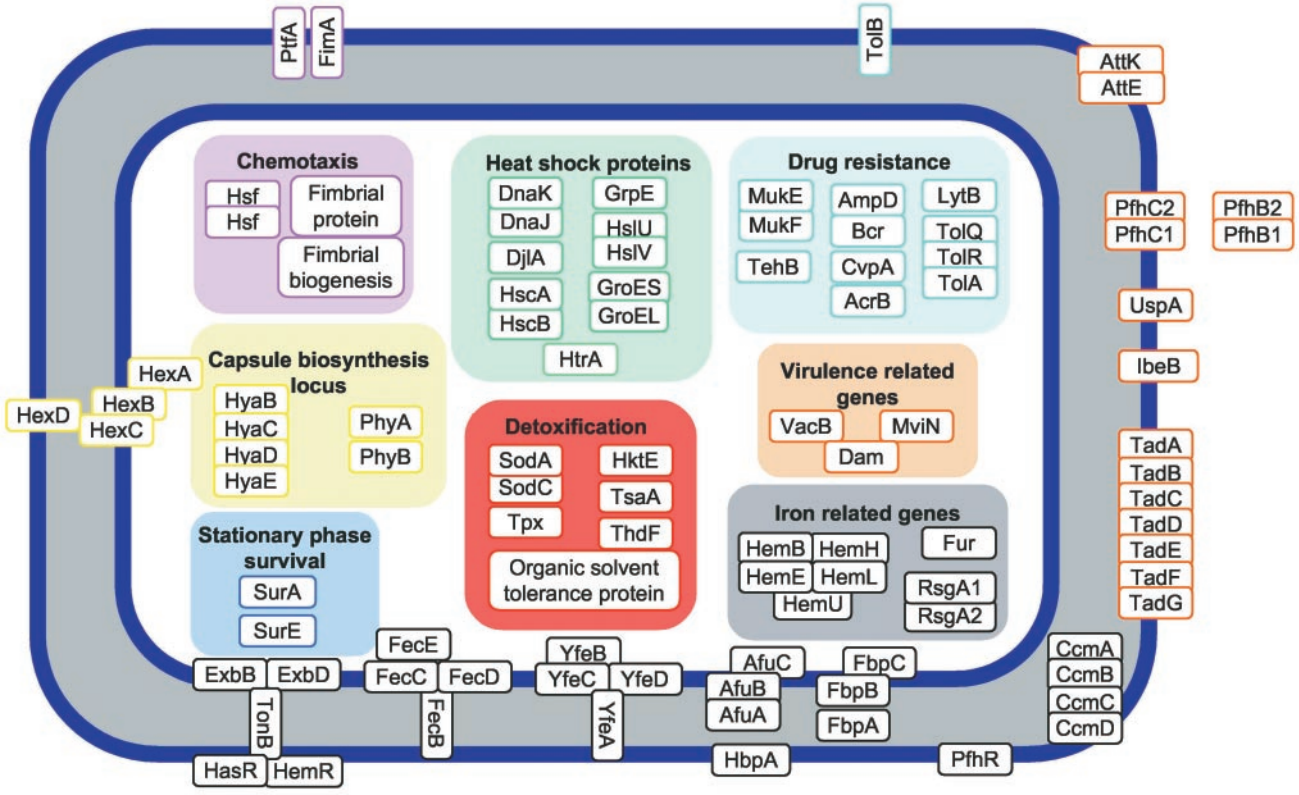

Fig. 4. A comprehensive view of the biochemical processes involved in *Pm* pathogenicity. Orthologs previously identified as virulence factors in other organisms are represented. Principal functional categories are shown in bold. Potential coding sequences related to these functions are arranged within each correspondingly colored category.

record for groups of vertebrates by plotting D versus time of divergence and obtained a slope of 0.088 that can be used to extrapolate evolutionary distance to time in millions of years. Applying this calibration to the 774 conserved genes, we estimate that *Pm* diverged from *Hi* ≈ 270 mya (range, 138–407 mya) and separated from a common ancestor with *Ec* ≈ 680 mya (range, 319–1,024 mya) (Fig. 3*b*). Together, these results place the root of the γ division of the proteobacteria at ≈680 mya, and confirm the relatively close relationship between the *Pasteurella* and *Haemophilus* genera (Fig. 3*b*).

Virulence Factors of *Pm*. The molecular basis for *Pm*'s pathogenicity and host specificity is not well understood. Our genomic analysis identified a total of 104 putative virulence-associated

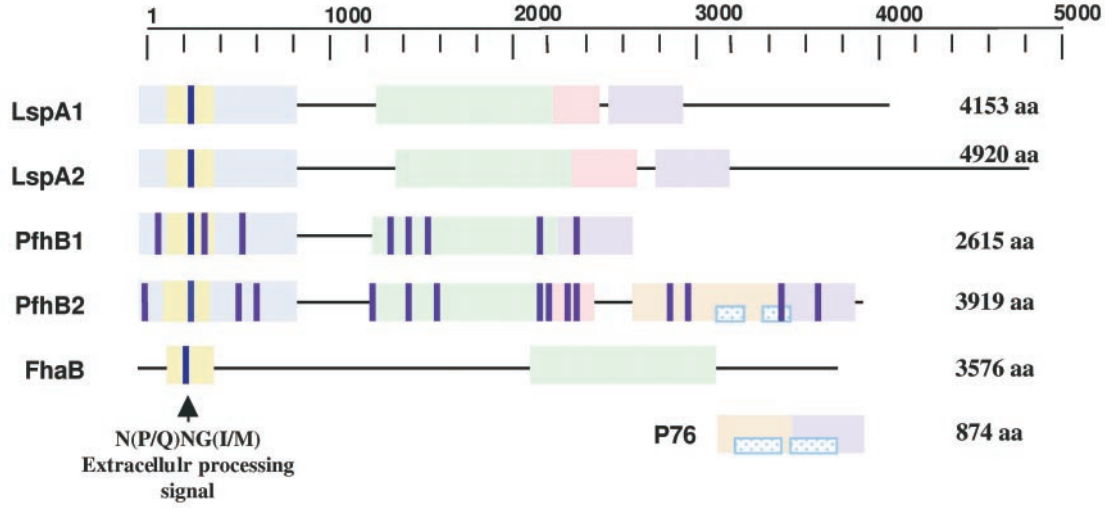

Fig. 5. Analysis of PfhB domains. Homology comparisons among PfhB1, PfhB2, from *Pm*, LspA1, LspA2, from *H. ducreyi*, FhaB from *B. pertussis*, and P76 from *Haemophilus somnus* are presented. Homologous domains are represented with the same colored boxes and the direct repeats in p76 and PfhA2 are patterned in blue. The N(P/Q)NG(I/M) extracellular processing motif is indicated and the integrin-binding protein motifs are shown as dark purple lines.

genes in *Pm*, which account for ≈7% of the coding density of the genome (Fig. 4). In particular, we discovered a 24-kb region with two new potential *Pm* virulence factors and their accessory genes. The two potential virulence CDSs, here termed *Pasteurella f*ilamentous *h*emagglutinin (*pfh)B1* and *pfhB2*, are 7,845 and 11,757 base pairs, respectively, and are virtually identical in sequence with the exception of a large deletion in the central region of *pfhB1* (Fig. 5). Both PfhB1 and PfhB2 contain domains with strong homology to the filamentous hemagglutinin (FhaB) of *B. pertussis* (19–21). FhaB governs the adherence of *B. pertussis* to host cells and is a major component of the acellular vaccines used to protect humans against whooping cough (22). Orthologs of this large protein have also been recently described in several other pathogens including *Haemophilus ducreyi* (LspA1 and LspA2), *Neisseria meningitidis*, *Serratia marcescens*, *Proteus mirabilis*, and *Pseudomonas aeruginosa* (23–27).

Several lines of evidence suggest that PfhB1 and PfhB2 play a role in the virulence of *Pm.* First, the amino-terminal ends of PfhB1 and PfhB2 have a conserved motif N(P/Q)NG(I/M) that is present in all members of this large protein family and is involved in posttranslational processing and extracellular signaling (Fig. 5; ref. 28). Second, the central region contains several integrin-binding motifs that also are characteristic of this family (29). The presence of these conserved motifs suggests that, similar to *FhaB of B. pertussis*, both PfhB1 and PfhB2 are involved in adherence of bacterial cells to host cell surfaces. Third, the carboxy terminus contains regions with significant homology (66% amino acid identity) to the serum-resistance protein p76 of *Haemophilus somnus* (30). Expression of p76 confers resistance to opsonization in *H. somnus*, thereby enhancing pathogen survival in the host. Two approximately 400-aa direct repeats found in p76 are also present in PfhB2 (30). Interestingly, the carboxyl-terminal region of p76 shows homology to LspA1 and LspA2 of *H. ducreyi*, the causative agent of chancroid ulcers in humans (23). Taken together, these studies provide compelling evidence that the PfhB protein plays a major role in the virulence of *Pm* and that its inactivation may enable the development of well-defined live attenuated vaccines.

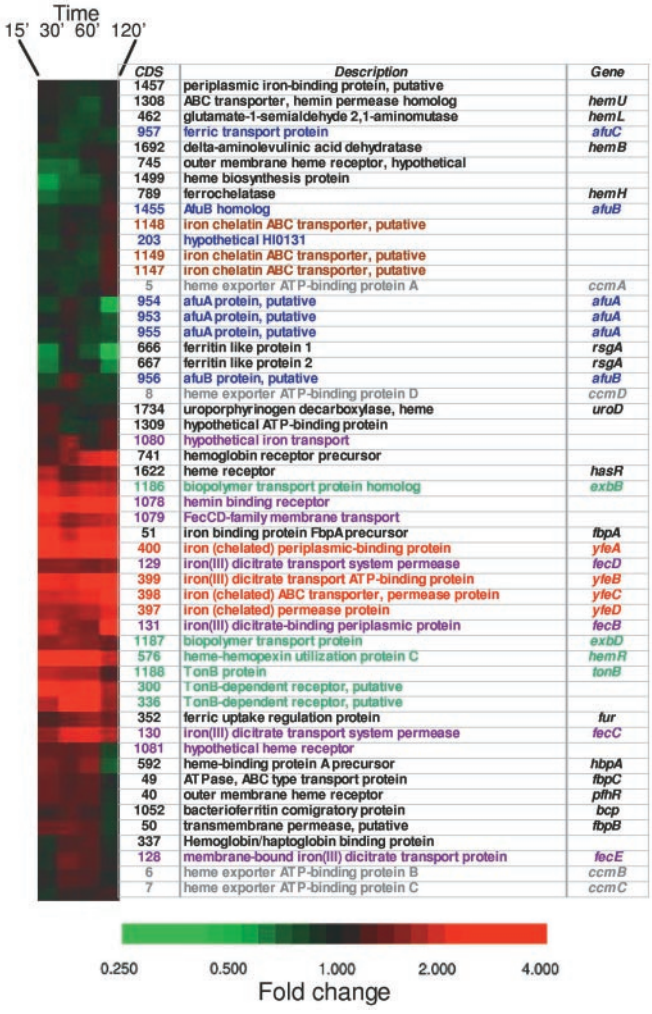

Fig. 6. Hierarchical clustering of 53 *Pm* CDSs identified based on sequence comparisons to be involved in iron metabolism or to require iron as a co-factor. Fluorescent ratios representing relative expression profiles of *Pm* CDSs at 15, 30, 60, and 120 min time points are shown in (*Left*). The red and green colors represent fold increase and decrease, respectively, in gene expression in response to low iron conditions (addition of the iron chelator, 2,2′-dipyridyl). Their corresponding open reading frames numbers and CDS descriptions are also shown. CDSs depicted in the same color are in close proximity to each other on the genome and may be a part of potential operons.

Iron Uptake and Acquisition in *Pm*. Because of its central role in electron transport, iron is an essential nutrient for most organisms. The low solubility of iron at neutral pH leads to a relative paucity of free iron within hosts and creates a major barrier to microbial growth *in vivo*. To compete for this limiting nutrient, microbes have evolved a variety of novel strategies for acquiring iron (31). More than 2.5% (53 CDSs) of the *Pm* genome is devoted to genes encoding proteins homologous to known proteins involved in iron uptake or acquisition (Fig. 6). Furthermore, several of these genes have undergone recent duplications (68–72% similar) in the *Pm* genome. For instance, *Pm* has a cluster of genes closely related to those encoding an iron transport system of *Actinobacillus pleuropneumoniae* (*afuA*, *afuB,* and *afuC*), but there are three copies of the *afuA* ortholog (Pm0953, 0954, and 0955; ref. 32) in the Pm70 genome. A second region (base pairs 520,397–528,389) of the *Pm* genome is rich in genes with homology to those linked to iron transport in the 102-kb pigmentation locus in *Yersinia pestis* (33). The 102-kb region of *Y. pestis* contains, for example, the *ybt* locus involved in siderophore uptake and the *hms* locus involved in hemin storage. Although the orthologs for *ybt* and *hms* are not found in strain Pm70, 8 other CDSs located between these genes in *Y. pestis* are present in the Pm70 genome. The predicted proteins of these eight CDSs contain transmembrane domains, and several are hypothesized to be members of the abc transporter family. One putative membrane CDS (Pm0446) contains a cytochrome *c* motif that provides a heme-binding site and another transmembrane CDS (Pm0453) contains a high-affinity iron permease motif. Because this region is not present in the *Ec* or *Hi* genomes, it has either been deleted in those genomes or acquired by *Pm* through horizontal transfer.

To elucidate other genetic systems involved in iron acquisition and metabolism, we used microarray technology to study patterns of gene expression in the *Pm* genome under iron depleted conditions. The results show that 174 of the CDSs showed a 2-fold or greater change in expression for at least one of four time points (15′, 30′, 60′, and 120′) after growth under iron-limiting conditions. Hierarchical clustering of these 174 CDSs on the basis of their expression patterns revealed that 53% decreased in expression in response to low-iron conditions, and 47% increased in expression relative to the control. Among these, reduced expression was seen for orthologs to the *frd* and *fdx* operons, both of which are involved in anaerobic energy metabolism and require iron to function (34, 35). In addition, 11 homologues involved in glycolysis or electron transport were decreased in expression under low-iron conditions. Of the CDSs that increased in expression, 16 were homologous to iron-specific cation transport proteins. Other orthologs in this group include

the surface polysaccharide biosynthesis enzymes GmhA and Skp. Thirty-one percent of the 174 CDSs whose expression was significantly altered by low-iron levels are previously undescribed or have close homology only to hypothetical proteins. The remaining 1,760 CDSs were either poorly measured or did not change significantly over the course of the experiment. Additional data, including the original hybridization images, a list of the genes represented on the array, as well as all primary datasets can be found at our web site (http://www.cbc.umn.edu/ResearchProjects/AGAC/Pm/Pmarraydata.html).

To specifically examine the behavior of known iron-related genes, we performed a second hierarchical cluster analysis focused on 53 CDSs with homology to iron-binding or uptake genes (Fig. 6). Several iron-related families of genes tended to cluster close to each other in terms of their expression profiles. Especially notable was the overall increase in expression of the *fec*, *yfe*, and *tonB* gene families. *Yfe* was previously identified in *Yersinia pestis* as an abc iron transport system (36), whereas *fec* and *tonB* were characterized as periplasmic-binding-protein-dependent iron transporters (37). Similarly, Pm0451 and 0452, which are orthologs to genes in the 102-kb pigmentation locus, were also up-regulated. In contrast, genes homologous to the *afu* family of periplasmic-binding-protein-dependent iron transport genes showed an overall decrease in expression (34). Taken together, these findings indicate that whole-genome expression studies made possible with the knowledge of the complete genome sequence provide a rich source of new hypotheses and insights for future research into the molecular basis of pathogenesis of *Pm*.

Concluding Statement. In summary, the entire genome sequence and preliminary functional analyses have identified a variety of virulence factors for possible investigation in this major animal and human pathogen. Furthermore, the availability of the genome sequence provides a foundation for future research into the epidemiology, evolution, and mechanisms of pathogenesis and host specificity of *Pm*.

We gratefully acknowledge Caroline Hoellrich and Alongkorn Amonsin for assistance with MLEE analysis. We appreciate the support of Judy Laber, Megan Lillihei, and Daniel Strom at the Advanced Genetics Analysis Center at the University of Minnesota, as well as Jun Yu at the University of Washington Genome Center (Seattle, WA) and his colleagues at the Beijing Human Genome Center (Beijing, China) for assistance during the sequencing phase of our project. We also gratefully acknowledge the support of Bruce Roe (Department of Chemistry and Biochemistry, University of Oklahoma, Norman, OK) and Ernest Retzel (Computational Biology Centers, University of Minnesota) for providing technical facilities and protocols during various phases of the project. This project was made possible by the continued encouragement of our colleagues, Larry Schook, Michael Murtaugh, Mitchell Abrahamsen, and Sagarika Kanjilal, to whom we are forever grateful. Funding for this project was provided by research grants from the Minnesota Turkey Growers Association, the Minnesota Agricultural Experimentation Center, the University of Minnesota Academic Health Center, and the U. S. Department of Agriculture's National Research Initiative (to V.K.) and by the National Institutes of Health (to T.S.W.).

1. Pasteur, L. (1998) in *Milestones in Microbiology: 1156 to 1940*, trans. and ed. Brock, T. D. (Am. Soc. Microbiol., Washington, DC), pp. 126–131.
2. Rhoades, K. R. & Rimler R. B. (1989) in *Pasteurella and Pasteurellosis*, eds. Adlam, C. & Rutter J. M. (Academic, San Diego, CA), pp. 95–114.
3. Rhoades, K. R. & Rimler, R. B. (1991) in *Diseases of Poultry*, eds. Calnek, B. W., Barnes, H. J., Beard, C. W., Reid W. M. & Yoder H. W. J. (Iowa State Univ. Press, Ames, IA) pp. 145–162.
4. Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986) *Appl. Environ. Microbiol.* **51,** 873–884.
5. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J.-F., Dougherty, B. A., Merrick, J. M., *et al.* (1995) *Science* **269,** 496–512.
6. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1999) in *Current Protocols in Molecular Biology* (Greene and Wiley–Interscience, New York) pp. 2.4.3–2.4.4.
7. Hunt, M. L., Ruffolo, C. G. & Rajakumar, K. (1998) *J. Bacteriol.* **180,** 6054–6058.
8. Frishman, D., Mironov, A., Mewes, H. W. & Gelfand, M. (1998) *Nucleic Acids Res.* **26,** 2941–2947.
9. Salzberg, S. L., Delcher, A. L., Kasif, S. & White, O. (1998) *Nucleic Acids Res.* **26,** 544–548.
10. Lowe, T. M. & Eddy, S. R. (1997) *Nucleic Acids Res.* **25,** 955–964.
11. Eisen, M. B. & Brown, P. O. (1999) *Methods Enzymol.* **303,** 179–205.
12. Salzberg, S. L., Salzberg, A. J., Kerlavage, A. R. & Tomb, J. F. (1998) *Gene* **217,** 56–67.
13. Lobry, J. R. (1996) *Mol. Biol. Evol.* **13,** 660–665.
14. Fraser, C. M., Eisen, J., Fleischmann, R. D., Ketchum, K. A. & Peterson, S. (2000) *Emerg. Infect. Dis.* **6,** 505–512.
15. Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., *et al.* (1997) *Science* **277,** 1453–1474.
16. Rimler, R. B. & Rhoades, K. R. (1989) in *Pasteurella and pasteurellosis*, eds. Adlam, C. & Rutter, J. M. (Academic, San Diego, CA) pp. 37–73.
17. Grishin, N. V. (1995) *J. Mol. Evol.* **41,** 675–679.
18. Feng, D. & Doolittle, R. F. (1997) *J. Mol. Evol.* **44,** 361–370.
19. Relman, D. A., Domenighini, M., Tuomanen, E., Rappuoli, R. & Falkow, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 2637–2641.
20. Locht, C., Bertin, P., Menozzi, F. D. & Renauld, G. (1993) *Mol. Micobiol.* **9,** 653–660.
21. Domenighini, M., Relman, D., Capiau, C., Falkow, S., Prugnola, A., Scarlato, V. & Rappuoli, R. M. (1990) *Mol. Microbiol.* **15,** 787–800.
22. Sato, H. & Sato, Y. (1999) *Biologicals* **27,** 61–69.
23. Ward, C. K., Lumbley, S. R., Latimer, J. L., Cope, L. D. & Hansen, E. J. (1998) *J. Bacteriol.* **180,** 6013–6022.
24. Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D., Chillingworth, T., *et al.* (2000) *Nature (London)* **404,** 502–506.
25. Poole, K., Schiebel, E. & Braun, V. (1988) *J. Bacteriol.* **170,** 3177–3188.
26. Uphoff, T. S. & Welch, R. A. (1990) *J. Bacteriol.* **172,** 1206–1216.
27. Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., *et al.* (2000) *Nature (London)* **406,** 959–964.
28. Braun, V., Hobbie, S. & Ondraczek, R. (1992) *FEMS Microbiol. Lett.* **79,** 299–305.
29. Ruoslahti, E. (1996) *Annu. Rev. Cell. Dev. Biol.* **12,** 697–715.
30. Cole, S. P., Guiney, D. G. & Corbeil, L. B. (1993) *J. Gen. Microbiol.* **139,** 2135–2143.
31. Wandersman, C. & Stojiljkovic, I. (2000) *Curr. Opin. Microbiol.* **3,** 215–220.
32. Chen, N., Frey, J., Chang, C. F. & Chang, Y. F. (1996) *FEMS Microbiol. Lett.* **143,** 1–6.
33. Buchrieser, C., Rusniok, C., Frangeul, L., Couve, E., Billault, A., Kunst, F., Carniel, E. & Glaser, P. (1999) *Infect. Immun.* **67,** 4851–4861.
34. Imlay, J. A. (1995) *J. Biol. Chem.* **270,** 19767–19777.
35. Ta, D. T. & Vickery, L. E. (1992) *J. Biol. Chem.* **26,** 11120–11125.
36. Bearden, S. W. & Perry, R. D. (1999) *Mol. Microbiol.* **32,** 403–414.
37. Moeck, G. S. & Coulton, J. W. (1998) *Mol. Microbiol.* **28,** 675–681.