Array-Based Genomic Comparative Hybridization Analysis of Field Strains of *Mycoplasma hyopneumoniae* \(\dagger\) †

Melissa L. Madsen, 1 Michael J. Oneal, 1 Stuart W. Gardner, 2 Erin L. Strait, 1 Dan Nettleton, 2 Eileen L. Thacker, 1 and F. Chris Minion 1*

Departments of Veterinary Microbiology and Preventive Medicine 1 and Statistics, 2 Iowa State University, Ames, Iowa

Received 6 July 2007/Accepted 4 September 2007

*Mycoplasma hyopneumoniae* is the causative agent of porcine enzootic pneumonia and a major factor in the porcine respiratory disease complex. A clear understanding of the mechanisms of pathogenesis does not exist, although it is clear that *M. hyopneumoniae* adheres to porcine ciliated epithelium by action of a protein called P97. Previous studies have shown variation in the gene encoding the P97 cilium adhesin in different strains of *M. hyopneumoniae*, but the extent of genetic variation among field strains across the genome is not known. Since *M. hyopneumoniae* is a worldwide problem, it is reasonable to expect that a wide range of genetic variability may exist given all of the different breeds and housing conditions. This variation may impact the overall virulence of a single strain. Using microarray technology, this study examined the potential variation of 14 field strains compared to strain 232, on which the array was based. Genomic DNA was obtained, amplified with TempliPhi, and labeled indirectly with Alexa dyes. After genomic hybridization, the arrays were scanned and data were analyzed using a linear statistical model. The results indicated that genetic variation could be detected in all 14 field strains but across different loci, suggesting that variation occurs throughout the genome. Fifty-nine percent of the variable loci were hypothetical genes. Twenty-two percent of the lipoprotein genes showed variation in at least one field strain. A permutation test identified a location in the *M. hyopneumoniae* genome where there is spatial clustering of variability between the field strains and strain 232.

Genetic variation is thought to occur among bacterial species as a survival mechanism in both adverse environmental and host niches. A natural consequence of evolutionary and environmental pressures, genetic variation in pathogenic species results in changing phenotypes. Genetic variation can occur by three general mechanisms, local nucleotide sequence changes, intragenomic recombination resulting in reshuffling of genome sequences, and acquisition of foreign DNA (1). It can also occur vertically and horizontally. Vertical transmission refers to passage of genetic material to siblings through cell division and its accompanying replication mistakes (point mutations, inversions, and spontaneous deletions). Horizontal transmission involves the acquisition of new genetic material. This could occur among closely related species by transformation of native DNA, by transduction via phages, or by conjugation mechanisms, giving rise to organisms with subtle changes in phenotype. Alternatively, it could occur between dissimilar species by similar mechanisms that result in dramatic changes in phenotype. For example, pathogenicity islands are thought to arise by uptake and insertion of large DNA segments that encode large blocks of genes related to a virulence phenotype (7, 21). While it is clear that numerous species of bacteria have acquired large segments of DNA (22), there is no evidence for gene acquisition in mycoplasmas.

Mycoplasmas are cell wall-less bacteria that are thought to be the smallest organisms capable of self-replication. Their genome sizes range from 580 kb to over 1,700 kb (18). Their small genomes do not restrict their ability to generate high rates of diversity, however. This type of variation is not a consequence of environmental signals but rather occurs through random events. There are numerous examples of both small sequence changes and recombination to introduce genetic variation in mycoplasmas (31). This variation is usually expressed by the generation of new chimeric surface molecules with high rates of antigenic diversity. The mechanisms by which this occurs include slipped strand mispairing during DNA replication and recombination between homologous sequences. Genetic variation can result in phase switching when it occurs within homopolymeric tracts of adenine in promoter regions (32) or in structural gene sequences (34) or by DNA inversion (13, 23). The generation of chimeric genes by intragenic recombination also occurs (12). There is no evidence that any of these mechanisms are operative in *Mycoplasma hyopneumoniae*, however. Analysis of the *M. hyopneumoniae* genome sequence failed to identify families of lipoprotein genes that could undergo phase switching through mechanisms employed by other mycoplasmas for surface variation (17).

*M. hyopneumoniae* is the primary agent of porcine enzootic pneumonia (19). There is increasing evidence that *M. hyopneumoniae* has a predisposing influence on other infectious agents (24, 25, 30). Genetic variation is known to occur in *M. hyopneumoniae* (2, 4, 9, 28), but there have been few studies examining the extent of variation within field isolates at the molecular level (27). Phenotypic variation does occur within *M. hyopneumoniae*, as described by Young and Ross in the context of protein immunoblotting (33), and in some cases within specific genomic regions (28), but no studies have been reported that examined genetic differences in field strains of *M. hyopneumoniae*.
within genes on a global basis. This is due to the difficulty in isolating and cloning M. hyopneumoniae from field samples and to the fact that adequate tools have not been available until recently (17).

The studies reported here examined genetic variation in M. hyopneumoniae on a genome-wide basis using microarray technology. The arrays were based upon the genome sequence of strain 232 (17). Our results for 14 field strains show that microarrays can be used to examine genetic diversity and that all of the strains of M. hyopneumoniae vary in at least one genetic locus.

MATERIALS AND METHODS

Mycoplasma strains and culture conditions. Pathogenic M. hyopneumoniae strain 232, a derivative of strain 11, was used in this study (16). Fourteen field strains were cultured from case studies from the Iowa State University Veterinary Diagnostic Laboratory (Ames, IA) and were from the United States Midwest. All M. hyopneumoniae strains were grown in Friis medium as previously described (16) and were from less than 15 in vitro passages. Cultures consisted of 125 ml of Friis medium in 250-ml Erlenmeyer flasks incubated at 37°C with slow agitation until the culture reached mid-log phase, as indicated by color change and turbidity. Mycoplasmas were pelleted by centrifugation at 24,000 × g, and the cell pellets were stored at −70°C until the chromosomal DNA was isolated.

Microarray. The M. hyopneumoniae microarray consists of PCR products (probes) spotted onto Corning UltraGAPS glass substrates (Corning, Inc., Big Flats, NY). Eighty-nine percent (620/698) of the open reading frames of strain 232 are represented on the array as PCR products that are approximately 125 to 350 bp long. Each product is a unique sequence even within paralogous families, as described by Minion et al. (17). No tRNA or rRNA sequences were included. The primer design, array construction, and validation methods used have been described previously (14, 15). Each slide was divided into two regions (upper and lower), and each region contained the full array of spots printed in triplicate in a noncontiguous well-spaced format. This design allowed two independent hybridizations simultaneously to reduce variation due to slide interactions.

Experimental design. TempliPhi-amplified DNA samples from field isolates were compared to control strain 232 DNA using a two-color experimental microarray design. Independent samples from one isolate labeled with one dye were paired with control samples labeled with an alternate dye; the samples were mixed and hybridized to the microarray. For 9 of the 14 isolates (95MP1501, 95MP1502, 95MP1503, 95MP1504, 95MP1508, 95MP1509, 97MP0001, 00MP1301, and 05MP2301), four independent field isolate DNA samples were paired with four independent DNA samples from control strain 232. In two of the four arrays, the control sample was labeled with Alexa 555 dye and compared to the field isolate sample labeled with Alexa 647 dye (Molecular Probes, Inc., Eugene, OR). The dye assigned to the control and treated samples was reversed for the other two arrays (dyerswap), and each region contained the full array of spots printed in triplicate in a noncontiguous well-spaced format. This design allowed two independent hybridizations simultaneously to reduce variation due to slide interactions.

NIH Image analysis. Images were analyzed to determine spots, and signal intensities were quantified using the softWorRx Tracker software package (Applied Precision, Inc.). Spot-specific mean signals were corrected for local background by subtracting spot-specific median background intensities. The natural logarithms of the background-corrected signals from a single scan were adjusted by using an additive constant so that all scans of the same array-dye combination had a common median. The median of the adjusted log background-corrected signals across multiple scans was then computed for each spot to obtain one value for each combination of spot, array, and dye channel. The data for the two dye channels on any given array were normalized using LOWESS normalization to adjust for intensity-dependent dye bias (29; http://www.stat.berkeley.edu/users/terry/sarray /lowess/in dex.html) and adjusted signal intensities were used. The intensity of each channel were adjusted by using an additive constant so that the median for any combination of array and dye was the same for all array-dye combinations. The difference in normalized signal for each spot was calculated by subtracting the signal intensity of the Alexa 647 dye from the signal intensity of the Alexa 555 dye. The differences for the triplicate spots were then averaged within each array to produce one normalized difference value for each of the 627 probe sequences.

Data analysis. A linear model of the difference in signal intensity for the two dyes was fitted for each probe sequence using the normalized data. The model included an overall mean for the difference in dye effect (Alexa 555 intensity minus Alexa 647 intensity) and, for each field isolate, a fixed effect for the difference in signal intensity between the control and the field isolate. As part of each linear model analysis, a one-sided t test for the difference in signal intensity being greater than zero was conducted for each probe. This test was chosen because in our experimental design signal intensities could only show a decrease, unlike RNA analyses, where values can show variation in both directions consistent with up- or down-regulation. The P values for all the probes and field isolates were then analyzed to obtained false discovery rates (q values) using the method proposed by Benjamini and Hochberg (3).

The analysis of the field isolate data suggests that certain locations of the genome may experience more variation across strains than would be expected by chance. A permutation test was employed to assess spatial clustering of the variation between field strains observed in regions of the M. hyopneumoniae genome (http://www.R-project.org). The test consisted of summing the number of field strains with significant variation from strain 232 in consecutively tested genes within a sliding window around the genome. A sliding window size of 10
consecutive tested genes was used. The data can be accessed through the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession number GSE8306.

RESULTS

TempliPhi reactions. Genomic DNA from all mycoplasma isolates was subjected to amplification by TempliPhi because of low chromosomal DNA yields for several of the field isolates. The optimal time of nebulization (15 min) and optimal nitrogen stream pressure (10 lb/in²) were determined empirically by taking samples at various time points during the shearing process and analyzing them by electrophoresis. All amplified DNA preparations were then sheared using this time point and this nitrogen pressure, and the fragment size was confirmed by electrophoresis prior to labeling.

To determine if any signal biases were introduced by TempliPhi during the amplification reaction, strain 232 chromosomal DNA was subjected to TempliPhi amplification, labeled, and compared to nonamplified DNA from the same DNA lot in a dye swap experiment. Equal amounts of amplified and nonamplified DNA were sheared and added to labeling reaction mixtures. The nonamplified chromosomal DNA sample was mixed with an amplified DNA sample with the alternate dye label and hybridized to one array; a dye swap mixture was hybridized to the array at the opposite end of the same substrate. The probe signal intensities were quantified, and the values after background subtraction were compared using correlation analysis. Since TempliPhi functions in a rolling circle mode and mycoplasma chromosomes are circular, no differences were expected between amplified and nonamplified samples, as confirmed by a Pearson’s correlation coefficient \( r \) of 0.9803 (Fig. 1).

Microarray studies. Data from each of the field isolate replicates were used in the statistical analysis. The statistical analysis indicated that 123 genes of the combined field isolates were significantly different from genes of control strain 232 at a \( P \) value of <0.004 and a \( q \) value of <0.20. The results are presented in Fig. 2 (exact locations are shown in the table in the supplemental material). The field strains differed in the number of genes that exhibited significant variation from control strain 232 genes. The strain with the most variation was strain 95MP1509, with 40 locus differences; strain 95MP1506 had 28 locus differences. One strain, 95MP1507, showed one locus difference at mhp606, while strains 95MP1505 and 97MP0001 showed only two differences. For 22 loci there were differences in more than one strain (Fig. 3). Of these 22 loci, 15

FIG. 1. Comparison of mean log signal intensity values of genomic DNA and TempliPhi-amplified chromosomal DNA. The data represent log mean intensity values following background subtraction. A correlation analysis was performed, and Pearson’s coefficient \( r \) was 0.9803.

FIG. 2. Scatter plot of genetic variation of field isolates. Positions at which there is genetic variation are shown for each M. hyopneumoniae field strain.

FIG. 3. Locations of locus variation in M. hyopneumoniae field strains. The loci that showed variation from strain 232 are shown. The locations of variable loci are indicated in the outer circle (●, variation in a single strain; ○, variation among two strains; △, variation among three strains; ■, variation among four or more strains). The locations of putative lipoproteins are shown in the inner circle (●, loci that show variation; □, lipoproteins that do not show variation). The hot spot of genetic variation is indicated by a gray bar.
were different in two strains, 2 were different in three strains, 1 was different in four strains, 2 were different in five strains, and 2 were different in six strains. Fifty-nine percent of the genes showing differences (72/123) were hypothetical with no known function. Twelve of the 51 lipoprotein genes showed variation (Fig. 3).

**Variation analysis.** The method used for identification of variation hot spots for *M. hyopneumoniae* field strains was derived from a permutation test with a sliding window design. Figure 3 shows the gene locations and number of field strains that showed significant variation from strain 232. Figure 4 shows the path of the sliding window across the genome with a window size of 10 genes. This path was determined by starting at “gene 1” and adding the number of field strains significantly different from strain 232 for each gene through “gene 10.” This resulted in a total of three significant variations. When the window was shifted one gene (“gene 2” to “gene 11”), there were also three significant variations. This sliding window continued around the genome, and since the genome was circular, the last genes were included in a window with the first genes. If the locations of significant variations were totally random, the sum in the window should vary up and down fairly regularly across the genome. To determine if there were hot spots of variation in the genome, 10,000 random permutations of the observed variation locations were examined using the statistical computing program R 2.4.1 (http://www.R-project.org). The window with the maximum sum was retained, and the 95th percentile of the 10,000 permutations was determined to be 15. Thus, an observed sum of 15 or greater would be expected to occur only 5% of the time by chance. The horizontal line in Fig. 4 shows this significance level at the 95% confidence level, and only one region of variation was declared significant. The genes in this region of variation are listed in Table 1.

### Table 1. Genes identified by hot spot analysis

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
<th>Instance</th>
<th>Region of variation</th>
</tr>
</thead>
<tbody>
<tr>
<td>mhp516</td>
<td>CH</td>
<td>1</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp517</td>
<td>UH, lipoprotein</td>
<td>0</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp518</td>
<td>UH</td>
<td>0</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp519</td>
<td>UH</td>
<td>0</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp520</td>
<td>pepF</td>
<td>0</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp522</td>
<td>UH</td>
<td>2</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp524</td>
<td>UH</td>
<td>2</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp525</td>
<td>UH</td>
<td>0</td>
<td>Strain J</td>
</tr>
<tr>
<td>mhp526</td>
<td>CH</td>
<td>6</td>
<td>76 (62)</td>
</tr>
<tr>
<td>mhp527</td>
<td>CH</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>mhp528</td>
<td>UH</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>mhp529</td>
<td>UH</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>mhp530</td>
<td>CH</td>
<td>1</td>
<td>-</td>
</tr>
<tr>
<td>mhp531</td>
<td>CH</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>mhp532</td>
<td>tsxE, lipoprotein</td>
<td>5</td>
<td>79 (12)</td>
</tr>
<tr>
<td>mhp533</td>
<td>CH</td>
<td>5</td>
<td>78 (12)</td>
</tr>
<tr>
<td>mhp535</td>
<td>UH, lipoprotein</td>
<td>2</td>
<td>92 (94)</td>
</tr>
<tr>
<td>mhp538</td>
<td>CH</td>
<td>1</td>
<td>97 (10)</td>
</tr>
<tr>
<td>mhp539</td>
<td>CH, lipoprotein</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>mhp540</td>
<td>tuf</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>mhp541</td>
<td>lon</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>mhp542</td>
<td>CH</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>mhp543</td>
<td>upp</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>

### DISCUSSION

Previous studies have shown that strains of *M. hyopneumoniae* vary in virulence potential (35) and that genetic variation does occur in this species, as measured by randomly amplified polymorphic DNA analysis (2, 27). To estimate genetic variation within *M. hyopneumoniae* in a more global, focused fashion, we performed comparative genomic hybridization on microarrays. This study utilized 14 field strains for comparison with virulent strain 232. Many low-passage field strains are difficult to propagate in vitro, a characteristic that has impacted the number of isolates available for analysis. In the United States, the swine serum component of the medium is highly variable in terms of its ability to support growth of recent isolates (E. L. Thacker, personal communication), which may explain the difficulty in isolating *M. hyopneumoniae* from clinical samples. The isolation of field strains is also often impeded by the more rapid outgrowth of other mycoplasma species in clinical samples, as well as the low growth rate of *M. hyopneumoniae*. To overcome low yields of chromosomal DNA for the slowly growing field strains, other methods of obtaining sufficient quantities of chromosomal DNA for the analysis were sought. TempliPhi is an enzyme that replicates DNA in a rolling circle replication fashion and was originally developed to amplify plasmid or viral DNA sequencing templates in lieu of culturing and template purification. Our preliminary studies indicated that TempliPhi was also capable of amplifying all regions of the AT-rich mycoplasma genome equally well without bias (Fig. 1). Additionally, since the field strain DNA and...
the control strain 232 DNA were both amplified by TempliPhi, any specific region bias would be equally reflected in the DNAs and thus not impact the analysis.

An initial analysis of the data showed that the field strains exhibited significant variation in their reactivities on the microarray (Fig. 2 and 3). Although most of the variation seemed to be randomly spaced around the genome, there seemed to be one hot spot of variation (Fig. 2). To test this possibility, the variation data were subjected to a permutation test with a sliding window size of 10 genes. Genes were considered individual units of the same size to simplify the analysis. The results of this analysis are shown in Fig. 4, where one region containing 23 genes was identified at a P value of <0.05. The genes within this region are listed in Table 1. When the sequences of these genes were compared by BLAST analysis with the two other M. hyopneumoniae published genome sequences (26), the sequences of strains J and 7448, it was apparent that the region encompassing genes mhp522 to mhp538 was missing from the genomes. There were an additional five genes in the region that were not part of the array, and three of these five genes, mhp521, mhp523, and mhp534, were also missing in strains J and 7448. Two genes, mhp536 and mhp537, were present in both strains. Thus, this region of the 232 genome is highly pleomorphic. Interestingly, all of the field strains in this study were from the United States Midwest and contained at least a portion of these sequences, as shown by our positive results for the microarray. This highlights one limitation of the analysis. Not all of the M. hyopneumoniae genes are represented on the array. The missing genes are listed in the table in the supplemental material. The genes missing in the variable region were not included in Table 1 since they were not included in the analysis.

One question of interest was whether lipoproteins showed variation among the different field strains. In other mycoplasma species, lipoproteins generate antigenic diversity as a consequence of phase switching and size variation (31). In M. hyopneumoniae, however, similar mechanisms of variation in surface proteins do not exist (17). Our results indicate that there is variation in lipoprotein genes in M. hyopneumoniae field strains since 12 of the 54 lipoprotein genes in the genome (17) varied among the field strains examined in this study (Fig. 3). Four of the lipoprotein genes (mhp517, mhp532, mhp535, and mhp539) were in the hot spot region of variation (Table 1). Only two of these genes showed variation; however, mhp535 varied in two strains, and mhp532 varied in five strains.

Interestingly, the P97 adhesin gene (8) varies in one strain (95MP1506), but its companion gene, P102 (10), does not seem to vary among field isolates. Both the P97 and P102 genes, however, have multiple paralogs in the chromosome (17). The P97 gene paralog mhp385 varied in strain 95MP1506, and mhp493 varied in strain 95MP1509. The P102 paralog gene mhp384 varied in two strains, 95MP1505 and 95MP1506, and mhp683 varied in strain 95MP1509. Although a limited number of field isolates were examined in this study, our data suggest that the cilium adhesin varies little in field isolates because of its critical role in adherence and colonization. The sequence of one of its paralogs, however, can vary, possibly as a way to introduce variation in the surface topography.

In unpublished studies, one locus was identified that displayed significant sequence variation in two of the field isolates used in this study, isolates 00MP1502 and 00MP1301 (E. L. Strait, M. L. Madsen, E. L. Thacker, and F. C. Minion, unpublished data). This chromosomal region involved mhp024 and included both a deletion and sequence variation. The region was identified using a nested PCR test that failed to identify these two strains with the inner primer pair (11). It is significant that the present study confirmed the variation within mhp024 in one of these strains (00MP1502), and the data are just outside our q value cutoff for 00MP1301 (P < 0.0053, q < 0.234) since the regions containing this sequence variation are represented on the array. Our analysis also showed genetic variation within mhp024 in strain 95MP1510.

These data indicate that the M. hyopneumoniae microarray can identify genetic variability among field isolates across the M. hyopneumoniae genome. A potential use of these results is to improve diagnostics by eliminating the variable genes from consideration for PCR targets. Ideally, the PCR target should be homogeneous across multiple field strains. In addition, the arrays can be used to screen other mycoplasma and bacterial species to enhance the specificity of the PCR target sequences for M. hyopneumoniae by eliminating the open reading frames that are cross-reactive. One limitation of this approach, however, is its inability to recognize DNA sequences present in field isolates but missing from the microarray. In summary, this microarray has proven to be a powerful tool for genomic analysis.

ACKNOWLEDGMENTS

We thank Nancy Upchurch and Barb Erickson for assistance with mycoplasma cultures. Monica Perez contributed to construction of the microarray. Mike Carruthers assisted with spot finding. Funding for this project was provided in part by the National Pork Board and the Iowa Livestock Health Advisory Council.

REFERENCES


Vol. 189, 2007

ANALYSIS OF FIELD STRAINS OF M. HYOPNEUMONIAE

7981


