

Kevin Dybvig

Regulation of a restriction and modification system via DNA inversion in *Mycoplasma pulmonis*

Kevin Dybvig* and Huilan Yu

Department of Comparative Medicine, University of Alabama at Birmingham, 418A Volker Hall, Birmingham, Alabama 35294, USA.

Summary

An invertible DNA element of 6.8 kb, designated the *hsd1* locus, was identified in the chromosome of *Mycoplasma pulmonis*. Infection of host cells with mycoplasma virus P1 revealed that the organism's restriction and modification (R-M) properties are controlled by inversion of *hsd1*. The nucleotide sequence of *hsd1* revealed several genes, the predicted amino acids of which bear striking similarity to the subunits of the type I R-M enzymes previously found only in enteric bacteria.

Introduction

The ability to respond rapidly to sudden changes in the environment is a fundamental property of organisms. Often, high-frequency phenotypic changes are mediated by DNA rearrangements. In bacterial systems, site-specific inversions are probably the best-studied examples of high-frequency DNA rearrangements. DNA-invertible elements usually act as on/off switches regulating expression of genes encoding surface organelles such as flagella, pili and fimbriae (Dybvig, 1993). As a consequence of elements such as these, diverse bacterial populations are maintained that have an increased ability to cope with environmental change.

Many mycoplasmas are human or animal pathogens, often causing chronic diseases of the respiratory tract, genital tract, and joints that can be difficult to eradicate. Little is known about factors that may contribute to disease pathogenesis, but lipoprotein surface antigens that are probably present in all species may be important factors (Wise *et al.*, 1992). In *Mycoplasma hyorhinis*, the species in which lipoproteins have been most thoroughly studied, these antigens have been shown to undergo rapid phase and size variation (Yogev *et al.*, 1991). Size variation results from a gain or loss in the number of repeats present in tandem in the gene encoding the protein moiety of the antigen. Other than the variable

lipoprotein antigens, genetic variation in mycoplasmas is largely an unexplored area. However, it has recently been shown that the chromosome of the rodent pathogen *Mycoplasma pulmonis* undergoes DNA rearrangements at a high frequency (Bhugra and Dybvig, 1992). Some of these rearrangements are attributable to transposition of the *M. pulmonis*-specific insertion sequence element IS1138 (Bhugra and Dybvig, 1993). However, IS1138 is involved in a minority of the rearrangements identified to date.

Previously, we reported that some DNA rearrangements in *M. pulmonis* correlate with the susceptibility of the cells to infection by mycoplasma virus P1, a virus thought to be specific for this species (Bhugra and Dybvig, 1992). Virus susceptibility in this earlier study was examined using an assay that was not quantitative. Using a more sensitive, quantitative assay based on plaque-forming units (pfu), we have discerned that changes in virus susceptibility correlating with DNA rearrangements are not an 'all or none' phenomenon. That is, cells are not actually virus sensitive versus virus resistant. Rather, as reported here, DNA rearrangements correlate with the cell's restriction and modification (R-M) properties, affecting the efficiency at which viral plaques are formed on lawns of susceptible cells.

The DNA fragments that undergo rearrangements correlating with changes in R-M have been fully cloned, revealing the presence of a 6.8 kb invertible element. The predicted amino acids of the major open reading frames (ORFs) within the element exhibit a high degree of homology with the type I R-M systems found in enteric bacteria, and DNA inversion acts as a genetic switch controlling the R-M properties of *M. pulmonis*. Type I R-M systems consist of three subunits: the R subunit is required for restriction but not modification, the M subunit is the methylase required for both restriction and modification, and the S subunit confers sequence specificity for DNA recognition and is required for both restriction and modification (reviewed by Bickle and Kruger, 1993). An analogue of each of these peptides is present in the predicted amino acid sequences encoded by the *M. pulmonis* invertible element. This DNA element is referred to as the *hsd1* locus.

Results

Identification and cloning of variable DNA fragments

Previously, restriction fragment-length polymorphisms (RFLPs) that were present in some subclones of *M.*

Received 29 October, 1993; accepted 18 February, 1994. *For correspondence. Tel. (205) 934 9327; Fax (205) 975 4418.

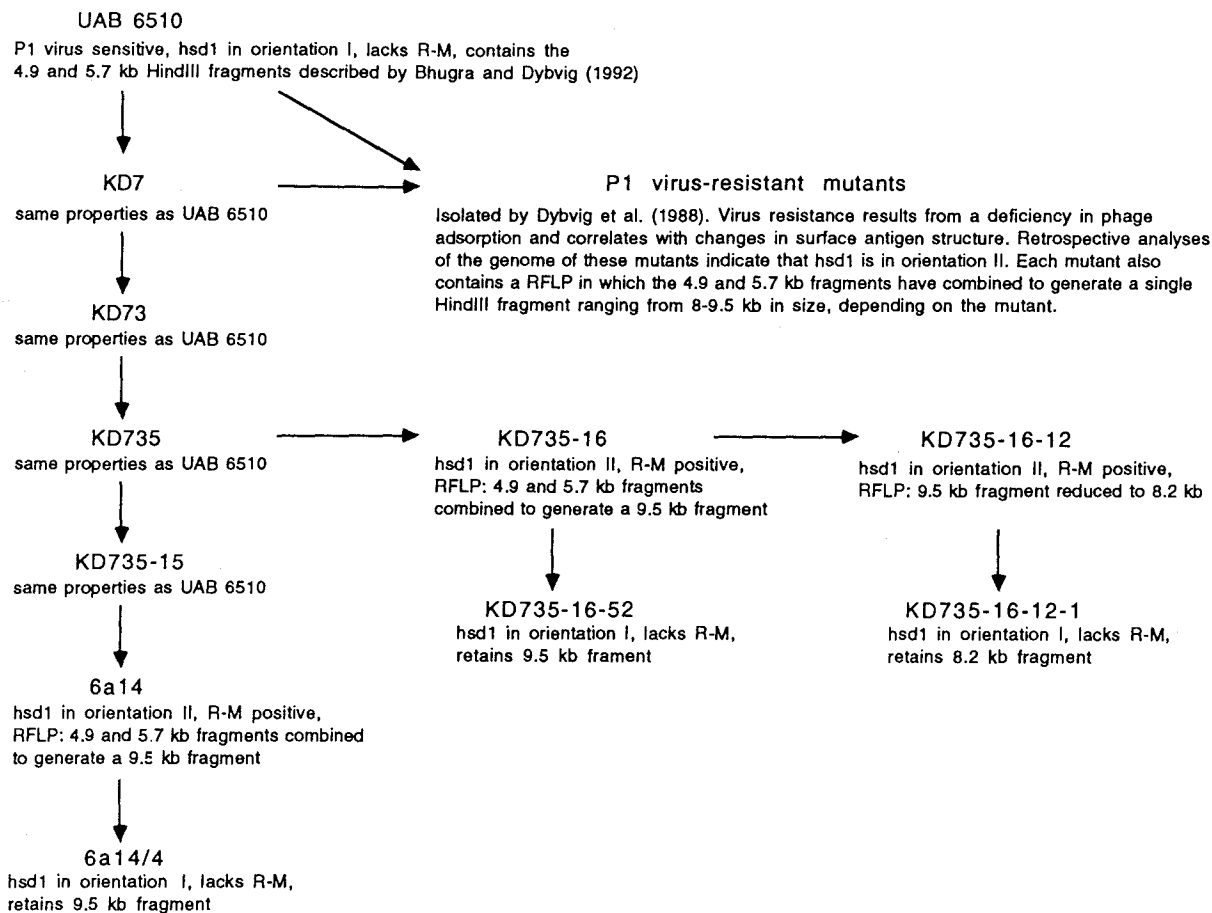


Fig. 1. Schematic diagram showing the relationship and pertinent properties of strains used in this study. The R-M properties of strains previously described by Dybvig *et al.* (1988) could not be assessed because these strains are resistant to P1 virus. The isolation of other strains, containing RFLPs, was previously described by Bhugra and Dybvig (1992). These strains were isolated by picking random colonies from agar plates and screened either for RFLPs (strains with the prefix KD) or for changes in sensitivity (variation in R-M) to P1 virus (strains 6a14 and 6a14/4).

pulmonis strain KD735 were described (Bhugra and Dybvig, 1992). The relationship of these strains to one another, as well as to mutants of *M. pulmonis* identified in an earlier study (Dybvig *et al.*, 1988) as being resistant to infection by mycoplasma virus P1 owing to failure of the virus to adsorb to host cells, is shown in Fig. 1. One of the RFLPs described by Bhugra and Dybvig (1992) involved *HindIII* fragments of 4.9 and 5.7 kb that recombine to generate a 9.5 kb fragment. This RFLP is not the subject of this paper, but it will be addressed later in the Discussion. Other RFLPs described by Bhugra and Dybvig (1992) were *HindIII* fragments of about 5.9 and 6.3 kb, identified in strains KD735-16 and 6a14. These fragments were found in concert, and from the next generation, subclones were obtainable that no longer had either of these fragments. In the present study, both of these fragments were cloned from KD735-16 and sequenced in their entirety, revealing their actual sizes to

be 6.2 and 6.6 kb. These fragments contain the *hsd1* locus, described below.

Southern hybridization analysis was used to identify the precursors to the 6.2 and 6.6 kb fragments, by probing DNA from the parent strain KD735. The 6.2 and the 6.6 kb fragments both hybridized to four *HindIII* fragments of 6.8, 6.0, 5.0, and 3.0 kb (Fig. 2, lane 1 of panels A and B). The 6.8 and 6.0 kb fragments contain the *hsd1* locus, illustrated in Fig. 3. The 5 kb and 3 kb fragments contain a second *hsd* system, designated *hsd2*, which has not been fully characterized but is highly homologous to *hsd1*. The 6.6 and 6.2 kb probes exhibit some cross-hybridization because of a 450 bp region common to both fragments (depicted in Fig. 3 as the region shaded in black). Therefore, when used to probe DNA from strain KD735-16, the 6.6 kb fragment (Fig. 2, lane 2 of panel A) and the 6.2 kb fragment (Fig. 2, lane 2 of panel B) hybridized most strongly to itself and hybridized weakly to one

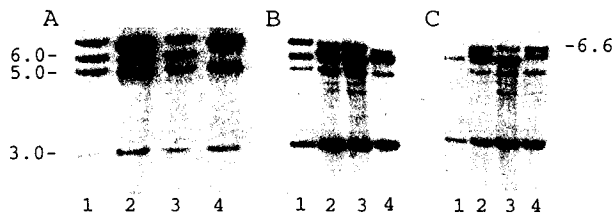


Fig. 2. Southern hybridization of *M. pulmonis* chromosomal DNAs probed with the cloned *Hind*III fragments of 6.6 and 6.2 kb from KD735-16 and 6.0 kb from KD735. Chromosomal DNAs were digested with *Hind*III, analysed by agarose gel electrophoresis, transferred to nitrocellulose filter paper, and probed with 32 P-labelled fragments.

A and B. Probed with the 6.6 kb fragment and the 6.2 kb fragment, respectively.

C. Probed with the 6.0 kb fragment.

In each panel, lane 1 contains DNA from KD735, lane 2 contains DNA from KD735-16, lane 3 contains DNA from KD735-16-12-1 (a derivative of KD735-16 in which the 6.2 and 6.6 kb fragments have reverted to their parental sizes of 6.0 and 6.8 kb), and lane 4 contains DNA from KD735-16-12 (a derivative of KD735-16 in which the 6.2 and 6.6 kb variant fragments have been retained). For reference, the top two bands in lanes 1 and 3 of (A) and (B) are the 6.0 and 6.8 kb fragments, and the top two fragments in lanes 2 and 4 of (C) are the 6.2 and 6.6 kb fragments. The 5.0 kb fragment is not observed in lane 1 of (C) because in this particular lane, the DNA was underloaded. Numbers in the margins refer to the size of molecular weight markers, in kb. The 6.8, 6.6, 6.2, and 6.0 kb fragments are diagrammed in Fig. 3, showing the degree with which these fragments should cross-hybridize.

another. The *hsd2* locus was also detected in KD735-16 DNA by using each of these probes.

In derivatives of KD735-16 and 6a14 in which the 6.2 and 6.6 kb fragments were no longer present, e.g. strain KD735-16-12-1, Southern hybridization data revealed that the 6.0 and 6.8 kb fragments were present again (Fig. 2, lane 3 of panels A and B). Therefore, a reversible DNA rearrangement apparently flips sequences back and forth between two states, first giving rise to the 6.2 and 6.6 kb fragments and then regenerating the 6.0 and 6.8 kb fragments. The 6.0 kb fragment was cloned from KD735 and, as expected, Southern blots probed with the 6.0 kb fragment revealed homology between this fragment and both the 6.2 and the 6.6 kb fragments from KD735-16 (Fig. 2, lane 2 of panel C). The 6.8 kb fragment from KD735 was not cloned in its entirety, although specific regions of this fragment were cloned and studied by enzymatic amplification of the pertinent sequences using the polymerase chain reaction (PCR).

Juxtaposition of the 6.2 and 6.6 kb fragments in the chromosome

The predicted amino acids encoded by an ORF located at one end of both the 6.2 and 6.6 kb fragments were strikingly similar to the peptide encoded by the *hsdR* gene of the *Salmonella* conjugative plasmid R124 (Price *et al.*,

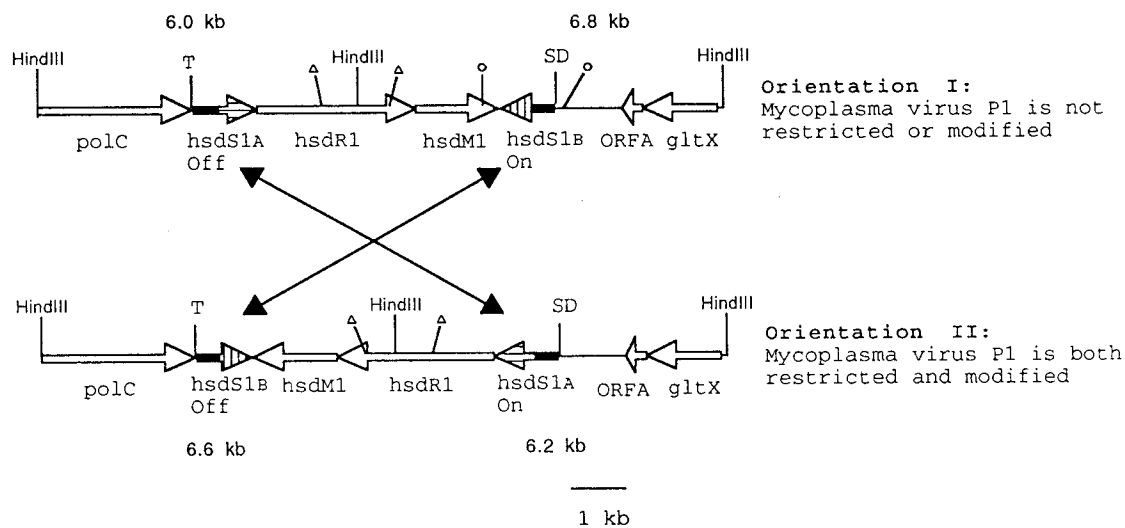


Fig. 3. Organization of the *M. pulmonis* *hsd1* locus. The putative gene products of ORFs, marked by arrows indicating directionality, were identified based on extensive predicted amino acid similarity with their counterparts in *E. coli* (*hsd* genes, GenBank accession number X13145) or *B. subtilis* (*gltX* and *polC*, GenBank accession numbers M55073 and X52116, respectively). DNA inversion apparently occurs within the 450-nucleotide region of sequence identity located at the 5' end of the *hsdS1A* and *hsdS1B* coding regions (shaded in black). The differing 3' halves of *hsdS1A* and *hsdS1B* are shown by horizontal and vertical stripes. The *hsdS1* gene in the 'on' position is preceded by a Shine-Dalgarno sequence (SD), and the *hsdS* gene in the 'off' position is preceded by a putative Rho-independent transcription terminator (T). The significance of the region located between *gltX* and the 5' end of *hsdS1* ('on' position) is unknown. Open circles indicate the target sites for primers used to PCR amplify *hsdS1B* from cells containing the *hsd1* locus in orientation I. Open triangles indicate the target sites for primers used to PCR amplify the junctions of the 6.2 and 6.6 kb fragments and the junctions of the 6.0 and 6.8 kb fragments.

AAGCTTTTAAAAAGTGGAGCAGATGTTTTAGTAGCTGGACATATTTAGTTTTTGGCCACAAAAGAAAAAATTTGCTAGCCTTTAGACTAGCTAAAAATTAATAAAAGGTTAAAAATGA 120
 AAATATGAAAAAACTAAGAACAGATATGCTCCTAGTCTACTGGATATTTGCACATTTGGAGGAGCTAGAACAGCTTTATTTAACTATTTATAGCAAAGCAGCTAATATGGTATTTAT 240
 GLIX-M K K L R T R Y A P S P T G Y L H I G G A R T A L F N Y L L A K H Y N G D F I

CATTGCTATAGAAGCACTGATGTCAAAGAAACATCGCTGATGGTGGGCTTCTCAAATGAAAAATTTAAAAATGATTAATAATAGAAGCTAATGAATCACCATTAAAACCAATGAAAA 360
 I R I E D T D V K R N I A D G E A S Q I E N L K W L N I E A N E S P L K P N E K

ATATGGTCTTTATAGACAAAGTCAAAGCTAGAAAAGTATTTAAAAATAGCTCATGAACATTAATGAAAAAGGCTATGCTTATAAAGCCTATGACAATAGTGAAGAACTTGAAGAGCAAAA 480
 Y G P Y R Q S Q K L E K Y L K I A H E L I E K G Y A Y K A Y D N S E E L E G E Q K

AAAACACTCTGAAAAATTAGGAGTAGCTTCTTTTATGATCAAAAGAGATTTTTAAAAATCTCTGAAGAAGAAAAACAAAAAAGAGACCGTAGCGGAGCCCTATTCTATTAGAGTAATTTG 600
 K H S E K L G V A S F R Y Q R D F L K I S E E E K Q K R D A S G A Y S I R V I C

CCCCAAAAATACCACTTACCAATGAGATGATTTAGTAAGGGGAAACATAGCAGTAAATTTCTAATGATATAGTGATGAAATAATCAAAAAAGTATGATTTATCCAACCTATAATTTTGC 720
 P K N T T Y Q W D D L V R G N I A V N S N D I G D W I I I K S D D Y P T Y N F A

AGTTGTAATGATGACATCGACATGAAATAGCCACATCTTGGCAGGAGAGAAGACACATACAAACACTCTAAACAAATGATGATCTATGATTTTAAATGCTCCAAAACCTTTGTT 840
 V V I D D I D M E I S H I L R G E E H I T N T P K Q M M I Y D Y L N A P K P L F

TGTCACCTTACAATTTACAATATGGAAGGCAAAAACCTTTCAAAAAGAGATTTATCTTTAAAACAATTTATTTATGAAATATAAAGAAGAGGATATAACTCAAGCTATTTTAA 960
 G H L T I I T N M E G K K L S A R G D L S L K Q F I H E Y K E E G Y N S Q A I F N

CTTCTTAACTTTGCTTGAATGAACTGATGAAAAAGCTCGAGAATTAATGGATCATGAAATTTAATTAATTTTATATACTAGATTTATCAAAATCTCCTTAAATTTGACATAAC 1080
 F L T L L G W T D E K A R E L M D H D E I I K S F L Y T R L S K S P S K F D I T

TAAAATGCAATGATTTTCTAAACAATATGGAAGGCAACACCTAATGAAGAGTTAATAAAAAATTTAAATTTAAATGACTATGATAATGATGAAATTAACCTATTTTATAGATTTATATA 1200
 K M Q W F S K Q Y W K N T P N E E L I K I L N L N D Y D N D W I N L F L D L Y K

AGAGAATATTTATTCATTAATCAATTAATAAAATTAATAAAATTTATAACAAGCAAAATTTAAATCAAGAAAAAGATCTTGATTTAAATGATGCAAAAAAATGATGATAATCATT 1320
 E N I Y S L N Q L K N Y L K I Y K Q A N L N Q E K D L D L N D A E K N V V K S F

TTGCTGATATTTGATTTCAAAATTTAGTGTAAATCAAAATCAAGAGGCGAATTAATAAACTCAAGAAAAATTTATCAATAAAGGTAATAATTTCTTCCCAATTAGAAAGGCAAC 1440
 S S Y I D I T S N F S N Q I Q E A I N K Q E K L S I K G K N L F L P I R K A C

TACCTTCCAAGACATGGTCCAGAACTTGTCTAAAGCTATTTATTTTGAAGTGAATTTATGAAAAAGAGAAATGAAAAATGAAATTAAGTAACTCAGAATTTACTCAAGAAAGT 1560
 T F Q E H G P E L A K A I Y L F G S E I I E K R M K K W K>END GLIX
 ORFA-M K I E F N S E F T Q E G

CAAGAAAAAATATTAATTTGCTCTCCGGTTGAAATAGTGAATATGAAGTGCAACTGTTTATGAATTTAGTGAACCTTCAAAGAGCATCTTTCTAGAAATAGAAATAAATAGTAAT 1680
 Q E K N I N F V S P V E I S E Y E G A T V Y E F S E P S K E H L S R I E I N S N

TCCAATGAAGTTTAAATAACCGCATCTATGACAATGATATTAATAATTAAGAAAGTGAANAATGCAATAGCACTAGGTCAGGCAAAATTTCTTAAATCAGCTTTTAAACAAA 1800
 S N E V L I I T T D S M T M I L K L N E E V E N A I A L G Q K F L K S L N K

TTAGAACAATGATAGTGAAGTTTTTTTGGCTATGTTTTAAAGATACATCAGGCAATGAATGGCTTGTTTTTAAATAACTTTAAAAATTTTGAATAATTTAAAAAGCTAATTTAT 1920
 L E H N D S E S F F A Y V L K D T S G N E L A C F K I T L K I F E>END ORFA

Terminator?

GCTTTAAAAATATAATTAATAAAACAGCAAAATAGTTGCTTAATTAAGCAACTTTTAAATTTACTTTTTATAAAAAATTTACTTTTTAATCATTCCAATTTAATTTATCTTTTTGATAA 2040
 AACATCAAAAAACAATTCCTTGTCTTTATTTTAGTCATAGCAAAAAAGTTGCTACACATAAATTAATTCATCTTTTTTGTCTTGATCTTCTGATTAATTTGGAGTATTTTCTATTT 2160
 TGGAAAAATAATCTCCAAATCTAAATTTATCTGTAGATTTTATGTTACTTCTTAAATTTTGGAGCATAAATTTTGTATATTTTAAATTTTCTAGTTGCTTTAATTAATTTTAT 2280
 TTTTGTATGCATCTACTTTTACAATTAAGATTAATAAATTTATCTAGATGTGCAAAATTTTTCAGCTTCAAAATTTGCAATAATACTAAATTTTGATAATTAATTTGATATGTAATC 2400
 AAGGCTATCTATATTGCTATTAAACAATCAAAACCTTCTATTCTTAAAGAAATGATTAATAATTTTTTTCTTTTTTCAAATTAATAAGAAATCAGATATGATAAATTTTGGCTTTT 2520
 TTAATGATCTTTTATGAATGAACTAATCTATTTTCTTTAGCATTGATTTACTTATCAATTAATAAAATAATAGAACAAATAACTTTTGGCTTTTGGATAATATAGATTTAAAT 2640
 ACCCTTTGTTTTAATGAAATCTTTTTCATGAGATAAATTTGAAGTAAATTTTGTAAATGATTTGCAAGTTTATAGTGGATTCATAGTTGAAATTTGCTTATATTTCTTTTAT 2760
 TTTTGTGTAATAAATAAGTTTACTAAAAAATTTATATTTTAGTATTTCTTTCTTGTATTTAATTTGTATTTAGATTTAATTAATAAACAACATCCAGTATTTGAATATAGAT 2880
 TTGCAAGATTTAATTAATAAATCAAAATTAATTTAAATTTCTTGAAGAAATTTAAATAATGATTTTGTATTTAATTAATAAATAAATTTACTTACAAAAATAGACATAGACTT 3000
 -35 promoter? -10

TAAACAGAAAAAGTGTCTATTTTTTTGCAATTAGTGTTTTTAAGCATAAAAATTTAAATTTATGTAATTTGATTTAGATGTTTATAATTAAGTGTTTAAGCTCTTATAACTTTATT 3120
 RBS

TGAAAACTAATAATTTTTTTTATTATTAAGAAAGCTTCACTAATGGAATTTATAAACTTGGTCAGATTTAAATTTAGAAAAAGGCAAAATCAAAATATAATGCAAAATATGTTTCTCA 3240
 HsdS1A-M E I Y K L G Q I L N L E K G K S K Y N A K Y V S Q

AAATATCGAATTTATAATTTTACTCTTCAAAAAACAAGATCAAGGTATTTTGGAAAAATCAATTCATATGACTTTAATGTTAATTTAATTTACCCTCAAGTGGTCAATATG 3360
 N I G I Y N L Y S S K T K D Q G I F G K I N S Y D F N G E Y I L I T T H G A Y A C

AGGAACAGTAAATATGTAATGAAAAATTTTCCACAACAAGTAAATGTTTTTCTTAAAGTTAATGAAAAATTTTAAAGCAAAATTTTAAAGTTATTTATTTGTTACAAGAAAA 3480
 G T V K Y V N E K F S T T S N C F I L K V N E N I V K T K F L S Y L L L L Q E K

DR

AACATTCAATGATATGCTATAGCTTCTCATATGTTTATTTAAAAAATATAACAATAATGATTTTGAAGTTAATTTACCTAACTTAAAAATCAAAATGCAATTAATTTAGAAATTAATG 3600
 T P N D M A I G S A Y G Y L K N Y N I N D F E V N L P N L K I Q S A I I K I I E

DR

AGCTTAAAGAAAGTTATTTTTTAGGCATAAAAAATCTGTGAAGAATGATAGTGAAGAAAAACAAAAAAGATTTAAGTATATTAATTAAGATTTATGAACTTTTGAAGAAACAAATAAA 3720
 P K E D L F F R H K N L V R I D S E E N T K K D L S I L I K I I E P L E K Q I N

TGCATTTGATGAAGTATTTGAGTGAGCAAAAAAGTCTCCACATTTTAAATTTATTTTTAAATAAAGTGCATCAATTAATCTTCAATTTTCAAAAAATATAAACTTGGTGAAT 3840
 A F D E L I L S E Q K S L Q H Y L N Y F L N K L A S I N P S I F K N Y K L G E I

AGCAAAAATATTAAGTGGTAAAACTCCTTCCACTGCAAAAAAGGAATCAAAAAAAGGAATACCTTTTTTGGTCCGGGAGATCTTGATAATATGGTCCAAAAAGATTTATTTACTTT 3960
 A K I L S G K T P S T A K K E L W K K E I P F P G P G D L D N M V P K R F I T F

TAATGAAAAATGATAAAGACTCTGGCACCATTTTATTTTCTCTGCAACAATTTGGGAAAGTGGGATTTTAGACAATTTATCTTAAATTAACCAACAAATAACTTCAATGAGG 4080
 N E K M I K R S G T I L F S S A A T I G K V G I L D N L S W F N Q I T S I E A

AAATAAATCAATGTTATGGATAAGTTTTTATTTTCTTCTTAAAAAATAAGTTCAAAAAATAAATTTGAAAAATCAAGCGGAACAATTTTCTTCAATTAATAAAAAATTTTGA 4200
 N N N Y V M D K F L F F L L K K I S S K I K F E N S S G T I F P T I K K K Y F E

AAATTTTACACTAGAAATTCCTAATCTAAAACTCAAAGTGCATATTAGGTATTATTGAACCATTCGATAAAAAAATTAATCTTTTAAAAACAAAAGAAAAATTCGTTGAAAAAGATT 4320
N F T L E I P N L K T Q S A I L G I I E P L H K K I N L L K Q K K L L E K R F

TATATACTATCAAAATCACTTAATCAAGGAGAAAATCAAAGATGAGTAGAAATGAACTCTTTATGAAAAAGAATTTGTAGATGACTGGTTAAAAATCAAAAATATGTAAAATTAGATA 4440
I Y Y Q N H L I K E K I K D E>END HsdS1A
HsdR1-M S R N E L L Y E K E F V D D L V K N Q K Y V K L D

TTAAAAATGAAGAGAAAAATTTGAATTAATTTTCGAAAAATATAGGTAGACTCAATAACATAGAATTAACCTCAACAAAAATTTTGGATATAAGAAGAGAATTAATTTCAACAATTCG 4560
I K N E E K I F E L I F E N I G R L N N I E L T Q Q N I L D I R R E L L S N N S

CATCTCTTATCGATTTGCTCAGTATTATGAGGTTTGTACTGTTAAAATTTATAAAAATGATGGTCCAAAAAATACGCCCTAAAGTTCGTTGATGAGAAAATGAGCTAATAATG 4680
A S S Y R F A Q Y L W G F D T V K I Y K N D G P K K I R L K F V D W E N W A N N

AAATCTATGTTTACAACAATTTCTACAGAGAGCGAAAAAACAATGGGTAAGAATTTGATTCGCTGATCTTAATCAATGGATTTCCCTTAATACTTTTGAATTTAAAGACAAA 4800
E F Y V L Q N P S N L G A K K N M G K R F D S L I L I N G F P L I L F E F K D K

GTGAAAATTAATAAAGCAATAAATCAATAGATGAATCATATAGAGGTAGTGTTTTAAATAAAGGATTTTATAGATTTATCAAAATTTGATAGGTTCAAAATTTGAGAAGTTAAGT 4920
S E N I N K A I N Q I D E S Y R G S V L N K G I F R F I Q I L I G S N F E E V K

TTCTAGCAAAATAACAAAAGAACAATAATAAAGATTTCTTCAATCAAAATGAACTCCGAAAAATGGTGGTCTTCAAAAAGAGCTCATAAAAAGATTTTAAACAAAATGCTTTAGAAG 5040
F L A N N K R T N N N K I L S F K W T S E N G G S S K E L I K D F L N K N A L E

AAATCTAAAAAATATGATGTTTTCAGAGATCAAAAGATGATGAAAAATTTTACTTAGACCATATCAACAAAGGCAATCAAAAAGCAATTAACCTTGTGAAAAACAATA 5160
E Y L K N Y V V F Q R S K D D E K I I L L R P Y Q Q R A I K K A I N F V E K Q L

AACTAATTTAGATGCTAAGCATAATTTAAACAATGCATATATGGCATAACAAGAGTGGTAAAACCTTTAACTTCATATAAAAATAGCTGAAATATTTCAAAAATAGTGATATTG 5280
K T N L D D A K H N L N N A Y I W H T T A G S G K T L T S Y K I A E I L S K N S D I

ATCATGTTGTTTGTAGTTGATCGTAATGATTTAAATGATCAAAACATCTCAAAACATTTCAAAAACCTTATGTCAAGTCTTAAAAATGAAAAATGATTTTAAATCAAGACACTTCAA 5400
D H V V F L V D R N D L N D Q T S Q T F Q K L M S S S K N E K I D F L N Q D T S

AGGACTTATATGAGATTTTCTTAAAAAGAAAAATTAATATAACTACTATTCAAAAATTAATAATATTTTATCATCTTATAAAAATGAAAAATGAAATTTTAACTAATAAAAAAT 5520
K D L Y E I F L K K E K L I I T T I Q K L N N I L S S Y K N E K I E F L T N K K

TTGTTTTATTATGATGAGTGCATCGTTCTAATGCTGGGTTAATGGGAAAAGAATTAAGATTTTGAATAATAGCATCATGATAGGTTTTTCAGGAACCTCAATATTTGAAGAAA 5640
F V F I I D E C H R S N A G L M G K R I K D F L N N S I M I G F S G T P I F E E

ATAATGATAGAGAACTCAAAAATCTTCGAAATGAAATGACTCATATAACATGAAAGATGCAATTTAGATAAAAATGTTCTTGGATTTAAAGTAGTAAATTAACCAAGAACAC 5760
N N D R E T Q K I F G N E I D S Y N M K D A I L D K N V L G F K V V N Y Y Q E T

GAATTTTTCGAGAAAATAACAATCTAATCTAGGAAAAATTAAGTCAATAAATTAATGTTTCAAAAGTAAACATTTAGATTTTCAAAATAATAGAACTACAATTAATAAGCTTTTG 5880
R I F R E N N S N L G A K I K S I I N V I K S K H L D F T N N R N Y N S I I A F

ATACAATCCAAGATGCTTTGACTTCTATGATGAATTTTATAAAAATGGATGTCGGTGATTTTTTGGCACTCCAATATTTAGTAGCTATTCAAAATGAAGAAAAAATGAAAAGTTTTTCA 6000
D T I Q D A L T F Y D E F Y K M D V G D I P A T P I F S S Y S N E E K N E K F F

ATTTAAAAGAACATAAAGAAAAATTTTAAAAAGATATGAAGAAAAATTAATACTCTCTTAAAGTTGAAGATTTTGACAAATATGTCATGATGTTCAATGACGCTTTTAAAGATATA 6120
N L K E H K E K I L K R Y E E K F N T S F K V E D F D K Y V N D V Q W R F K E Y

ATAGTAAAAATAATAGCATTGATATAGTTATTGATGCGATATGTTAATACAGGTTTTGACTCACCAAGAACCAACTTTATATATAAACAAGAATTAAAAATCATAATTAATTC 6240
N S E N N S I D I V I V V D M L L T G F D S P R T N T L Y I N K E L K N H N L I

HindIII
AAGCTTTTCTAGAACAAATAGATTAAGTATTACTCAAAAAGCGTGAATAATTTGTTAATTTTCTCTCTGAAGAACAATCAATCAATGATGCTTTTAAAAATATATGCAAAATAGTAGT 6360
Q A F S R T N R L S D Y S K K R G I I V N F S L E E Q S I N D A F K I Y A N S S

ATAAAGAATACAAACTTGTATTGAGGAAAAATATGAACAGGTAGTTGAAGATTTTATTAATTTTGAATAGTTTAAAAATCTTTTTCTAAAAATTTATGATGAGAAAAATAATG 6480
D K E I Q Q L V Y G E K Y E Q V V E D F I N F W N S L K I S F S N I Y D E K N N

AGATTTTGAAGAAATTTCTTTAGAAAATAAAGAAAAATTTTAAAAATCTTAGTCAAGTTTCAAAATTTTCTTCTTAAAAACTTTTAAAGAATATGAAAAAATGAAAAATTT 6600
E I F R N I S L E N K K Y L K N L S Q V S N I F S S L K T F K E Y G K N E K I

CAGATTTTCTTTGGAACAATTAATCAATATCAAAAATGAGCAAAATGAGATCAAAAATTTATCTACTAATGAAAAAGAAAAATAAGTTATGAAGTTTTAAACTCAATTGATATTA 6720
S D F S L E Q L N Q Y Q K W A N E I K K N L S T N E K E K I S Y E V L N S I D I

GCAATATTAATTTGCTTTATAAAGAAATGATAATTGATGAAATATATTTGGAATTTTATTATCTTTAATAAAAAAATAGTAAATACCCAAATAATAGATTAACCTTATGAAGATACAT 6840
S N I K F A Y K E M I I D E I Y L E N L L P F N K K I S K Y P N N R L T Y E D T

TAAGTGAATAGATAAACAATTTCAATGATAAAAAAATAATACAATCAAGGAAAAATTAATCAAAAAGAATATGAATATTTCTTACTTTTAGTCAAAAATGAAAAATGAAATAAAAA 6960
L S E I D K H I Q L I K N N Y N Q G K I N Q K E Y E I F L L L V Q K W K N E I K

ACTTTTTATCAAAAAGATAAATCTTAGATGAAAAAGAAATTTATAGATTTGGAAAAAGGATTTTGAAGTGTTTTTCAAAAAGTTAAAAACCAAAATAGAAGCCATGTAOACTTGAAA 7080
R F P I K K D K S L D E K E F I D D Y G K R I L K S V F Q K V K N Q I E A M W L E

AAATCTTAAGAATATCATGGAATTAATAATGATCAAAATCGCAAGATTTAAAAAAGAATTAACGATAAAGATCTAGATGATTTGAAAAATCTGAATTTATAAAAAAATGATCAA 7200
K I L K E Y H G I N N D Q I R K D W K K R I N D K D L D D I E K S E F I K K W S

GAAGATCTAAAGAAGTTGATAAGGATATTATTGATAAATGTCATTAATAAAGAAAGCATTGAGGCCTTTTAGACTTTGAAATAAAAATGAATAAAAATATAGAAAAGCAAAATTT 7320
R R S K E V D K D I I D K L S I E Y K E S I E A F L D F E I K M N K I I E S K I >E
Mstart

AAAGTATGCAAAATAGTAAAGAATTAATAGCAGTAGTAAAAAATCTGTGATCAATTAAGATCAAAAATGGAAGTAACTGAATATAGAGATTACATAATGGGTTTTTGTTTTCAAT 7440
ND HsdR1
HsdM1-M S N S K E L I A V V K K I C D Q L R S K M E V T E Y R D Y I M G F L F F K

ATCTTTCAGACAATCTGAAAAAATTTGAAGAGTTTAAAGAAAGAGTTGATTATATTAATATCTGAAATTTGATGAAAATCATGAACAGTTTAAAAAATCAAGGAAATAATTTTC 7560
Y L S E Q S E K N F E E F K E R V D Y I K Y S E F D E N H E Q F K K I K E I I I

AAAAATGATGATTTTTTTCTAGCTTATAAATATAGTTTTCAAAATGTTGATGATGATGAATCAAGGAAAAATGTTATCTTCAATTAAGAATCTTTTAAATAAATGAAAGTA 7680
Q N D D D F F L A Y K Y S F Q N V V D M M N Q G K N V I P T I E E S F N K I E S

<i>M. pul.</i>MSRNELLYEKEFEVDDLKVNQKYVKLDIKNEEKIFELIFENIGRLNNIELTQQNI	54
<i>StyR124</i>	MTHQHTHTIAESNNFIVLDKYIKAEPDGSYQSESDL..ERELIQDL.RNQQYEFISVKSQSAMLANVREQLQNLNGVVFNDSEW	81
<i>M. pul.</i>	LDIRRELLSNNSSASSYRFAQYLWGFDTVKIYKNDGPKKIRLKFVDWENWANNEFYVLQQFPTRDAKNMGRFDSLLINGFPL	138
<i>StyR124</i>	RRFTEQYLDNPSDGLDKTRKIH.IDYICDFIFDDEERLENIYLIIDKKNLMRNKQVQIIQQF...EQAGSHANRYDVTILVNLPL	161
<i>M. pul.</i>	ILFEPKDKSENINKAINQIDESYRGSV.LNKGIFRFIQILIGSNFEVVKFLANNKRTNNKILSFKWTSSEGGSSKELIKDF..	219
<i>StyR124</i>	VQIELKRRGVAIREFANQIHRYSKESFNSENLSFKYLQLFVISNGTDRYFANTTKRDKN...SFDFTMNWAKSDNTLTKLDK	239
I		
<i>M. pul.</i>LNKNALEEYLNKYVVFQRSKDDEKIILLRYPQORAIKKAINFVEKQLKTNLDAKH...NLNNAYIWHHTGSGKTLTSYK	295
<i>StyR124</i>	FTATCFQKHTLLNLVNVYSVFDSS...QTLLVMRYPQIAATER...ILWKIKSSFTAKNWSKPESGGYIWHHTGSGKTLTSFK	319
IA		
<i>M. pul.</i>	IAEILSKNSDIDHVVFLVDRNDLNDQTSQTFQKLMSSSKN..EKIDFLNQDTSKDLVEIFLKKELIITTIQKLNILSSYKNE	377
<i>StyR124</i>	AARLATELDFIDKVFVVDKDLDYQTMKEYQRFSVSDSVNGSENTAGLKRNLDKD.....DNKIIVTTIQKLNLM...KAE	393
II III IV		
<i>M. pul.</i>	KIEFLTNKKFVFIIDECHRSNAGLMGKRIKDFLNNSIMIGFSGTPIFEEN..NDRETQKIFGNEIDSYNMKDAILDKNV LGFKV	458
<i>StyR124</i>	SDLPVYVQQVVFIFDECHRSQFGEAQKLNKKKFKRYQ QFGFTGTP IFPENALGSETTASVFGRELHSHYVITDAIRDEK VLFK .	476
<i>M. pul.</i>	VNY YQETRIFR...ENNNSNLG.....KIKSIINVI....KSKHLDFTNRRNYSIIAFDITQDALTFYDEFYKM	522
<i>StyR124</i>	VDY NDVRPQPKSLETETDEKKLSAAENQQAFLHPMRIQEITQYILNFRQKTRHTFPGSKGFNAMLAVSSVDAAKAYATFKRL	559
<i>M. pul.</i>	..DVGDI FAT.....PIFSSYSNEEKN.....EKF.....FNLKEHKEKILKRYEEKFNFSFKVE..DFDKYVNDVQWR	582
<i>StyR124</i>	QEAAANKSATYKPLRIATIFSFAANEEQNAIGEISDETPDTSAMDSSAKEFLDAAIREYNHFKTNFSTDSNGFQNYRDLAQR	644
V VI		
<i>M. pul.</i>	FKEYSENNSIDIVIVVDMMLLTGFDSPRNTLYINKELKNHNLIQAFSRTNRLSDYSKKRGIIVNFSLQESINDAFKIYANSS	666
<i>StyR124</i>	VK....NQDIDLLIVGMFLTGFDAPTLLNTLFVDKNLRYHGLMQAFSRTNR IYDATKTFGNIVTFRDLERSTIDAITLFG...	720
<i>M. pul.</i>	DKEIQQLVYGEKYEQVVEDFINFWS.LKISFSNIYDEKNEIFRNISLE...NKKKYLKNLSQVSNIFSSLKTFFKEYGKNEKI	746
<i>StyR124</i>	DKNTKNVVLEKSYTEYMEGFTDAATGEAKRGFMTVVSELEQRFPDPTSIESEKKEKDFVKLFGYELRAENILQNYDEPATLKAL	804
<i>M. pul.</i>	SDFSLEQLNQYQKWANEIKKNLSTNEK..EKISYEVLN SIDISNIKFAFKEMIIDEIYLENLLFFNKKISKYPNNRLTYEDTLS	828
<i>StyR124</i>	QQIDLSDPVAVEKFAE...HYVDEKFAELQTI RLPADRRIQDYSAYND.....IRDWQRREKEAEKKEKSTTDWDDVVF	878
<i>M. pul.</i>	EIDKHIQLIKNNYNQKINQKEYEIFLLLVQKWKNEIKNFKIKDKSLDEKEFIDYGRKILKSVFQKVNQIEAMWLEKILKEY	912
<i>StyR124</i>	EVD...LLKSQ...EINL.DYILGLIFEHNRQKNGKGEEMIEEVKRLIRSSLGNRAKEGLVVDFIQQTNLDDLDPKASIIDAF	953
<i>M. pul.</i>	HGINNDQIRKDWKRRINDKDLDDIEKSEFIKWSRRSKEVDKDI.....IDKLSIEYKESIEAFLDFEIKMNKIIIESKI*	987
<i>StyR124</i>	FTFAQRREQREAEALIKEENLNEDAARKYIRTSCLKREYATENGTELNETL PKLSPLNPQYKTKKQAVFRKSSRLRLSLKA*..	1034

Fig. 5. Alignment of the predicted amino acids of the *M. pulmonis* HsdR1 peptide with the R subunit of *StyR124I* (GenBank accession number P10486). Protein sequences were aligned using the GAP program. Highlighted in bold are the seven 'DEAD-box' regions conserved in the superfamily II of helicases (see text). (I), no substitution; (:), conservative substitution; (.), semi-conservative substitution.

ORFs present on the 6.2 and 6.6 kb fragments is shown in Fig. 3. The nucleotide sequence of the entire 6.0 kb fragment was determined, revealing that this fragment was a composite of the sequences of the 6.2 and 6.6 kb fragments as depicted in Fig 3. These data suggested that DNA inversion of a 6.8 kb element (the *hsd1* locus) was responsible for interconversion of the 6.0 and 6.8 kb fragments with the 6.2 and 6.6 kb fragments. Inversion apparently occurred between identical, inverted repeats of 450 nucleotides that flank the element and are located in the

5' halves of the ORFs designated *hsdS1a* and *hsdS1b*. To confirm that the DNA rearrangement was an inversion, oligonucleotide primers were used to PCR-amplify sequences flanking the *hsdS1b* gene from the parent strain KD735, containing the 6.8 kb fragment. As expected, the sequence of this PCR product revealed that this gene was in the chromosome in an inverted orientation relative to *hsdS1b* from strain KD735-16, indicating that the rearrangement was an inversion.

Inversion of *hsd1* apparently occurs at a high frequency,

<i>M. p.</i> S1A	-----MEIYK	LGQILNLEKG	KKYNAKYVS	QNIGIYNLYS	SKTKDQGFPG	KINSYDFNGE	YILITTHGA-	64
<i>M. p.</i> S1B	-----	-----	-----	-----	-----	-----	-----	64
<i>EcoDXXI</i>	MSELSYLEKL	LDGVEV·WVT	··SMADIGTG	S·---NRQDE	SEN··PF·V	RSKN·---IL	·SDTFE·DEV	A·V·PGE·GI 73
<i>StyR124</i>	MSEMSYLEKL	LDGVEV·WLP	··E·TKY·QP	TKYLVKAKDY	HDTYTIPVLT	AGKT·---FI	LGYTNETH·I	·QASKAPVII 76
<i>M. p.</i> S1A	---YAGTVKY	VNEKPFSTSN	CFILKVNENI	VKT--KFLSY	LLLLQEKTFN	-DMAIGSAYG	YLNKNYNINF	EVNLP----- 133
<i>M. p.</i> S1B	-----	-----	-----	-----	-----	-----	-----	133
<i>EcoDXXI</i>	GDIF·---H·	·EG·YALHQR	AYRIRITT·A	·D·---·Y·	FMSSSF·QYI	LTKSV·ATAI	SIRKPMLEG·	K·PI·SPDNP 147
<i>StyR124</i>	FDDFTTAN·W	·DFD·KAK·S	AMKMTSCDD	N·LL·YVY·	W·NTLPSE·-	---AEGDHK	RQWIS·YSQK	KIPIPCDNP 151
<i>M. p.</i> S1A	--NLKIQSAI	<u>IKLIEPKEDL</u>	FFRHKNLVRI	DSEENTKKDL	<u>SILIKIEPL</u>	EKQINAFDEL	ILSEQKSLQH	YLNFLNKLKLA 211
<i>M. p.</i> S1B	-----	-----	-----	-----	-----	-----	-----	·····FG·FY 178
<i>EcoDXXI</i>	EKS·A···E·	VR·LDT---	-----	-----	-----	-F	TAL <u>TABLTA</u> E	LTA·LNMRRK
<i>StyR124</i>	EKS·A···E·	VR·LDT---	-----	-----	-----	-F	TAL <u>TABLTA</u> E	LTA·LNMRRK
								QY·YRDQ·L 194
								QY·YRDQ·L 198
								***** +
<i>M. p.</i> S1A	SINPSIFKNY	KLGEIAKILS	<u>GKTPSTAKKE</u>	LWKKE--IPF	FGPGDLDMV	PKRFITFNEK	-----MIK	RSGTILFSSA 282
<i>M. p.</i> S1B	Q·E··L·HD·	··EK···RR	··IINSFDLK	ENPGD--Y·V	ISSNTKN·GI	----FGYLSN	-----YMY	DGEY·TI·AD 245
<i>EcoDXXI</i>	·FKEGEVEWK	T····GNFTY	·YAAKAMD--	--SGDARFVR	ITDINK·GKL	S·ENPMYV·L	NEENEKYTLD	KNDLLMARTG 270
<i>StyR124</i>	·FKEGEVEWK	T····GKW·Y	·GGTPSKN·I	EFWENG·S·W	IS·K·MGRTL	VDSESDYI--	--TEEAVLHS	STKL·PAN·I 273
<i>M. p.</i> S1A	ATIGKVGILD	NL-----SW	FNQQTISIEA	NNYV--MDK	FLFPLKKIS	SKIKF----E	NSSGTIFPTI	KKYFENFTL 350
<i>M. p.</i> S1B	GAYAGTVFLN	·G-----KF	SITNVCFL·LL	L·DK·NLLT·	···YY··NE	NI·QK----K	SIV·SSR·SV	REYTLSEIAI 315
<i>EcoDXXI</i>	·F··TM·FE	EDYPAVYAGF	LKLN-----	-L·ETIINA·	YYWHFA---Q	·DFFWEQANK	LV·GGQ·QF	NANALKQVRV 341
<i>StyR124</i>	·IVVRSS·D	KVLPALIKV	PATLNQDMKA	VIPHENILV·	YIYHMIGSRG	·D·LR--AAK	KTG·SVASID	S·NYPHLKI 351
<i>M. p.</i> S1A	EIP-----	NLKQTSAILG	IIEPLHKKIN	LLKQKKKLE	KRFIYYQNH	IKEKIKDE--	-----	----- 401
<i>M. p.</i> S1B	K·-----	S·EI····	·N·HF·YV--	-----	-----	-----	-----	----- 336
<i>EcoDXXI</i>	P·YPSHPQK	S·DE·GR·VD	·LDKFDIAIA	SITE--G·P	REIELR·KQ-	-----Y·YY	RDLLFSFPKP	ETVSN 406
<i>StyR124</i>	PV·-----	·INE·QR·VE	·LDKFDTLT·	SITE--VFR	VKSSCAR·NT	STIV·YCSVS	RNLKLSVIN-	----- 410

Fig. 7. Alignment of the predicted amino acids of the *M. pulmonis* HsdS1A and HsdS1B peptides with the S subunits of *EcoDXXI* (GenBank accession number X73984) and *StyR124I* (GenBank accession number X13145). Protein sequences were aligned using the PILEUP program. Dots refer to amino acids that are identical to the corresponding amino acid in the HsdS1A sequence, and dashes refer to gaps introduced into the sequences to improve alignment. Underlined is the 6-amino-acid direct repeat flanking the 33-amino-acid region present in HsdS1A but lacking in HsdS1B. The TAEI repeat region characteristic of type IC S subunits is denoted by plus signs, and the motif characteristic of ATP/GTP-binding proteins (see text) present in HsdS1A is denoted by asterisks.

these proteins are most closely similar to the S subunit of the R-M system encoded by *Escherichia coli* plasmid pDXXI (GenBank accession number X73984). The *EcoDXXI* enzyme shares significant homology with *StyR124I*; both of these are classified as type IC enzymes (Bickle and Kruger, 1993). HsdS1A and the S subunit of *EcoDXXI* share 23% identity and 49% similarity, and a comparison of HsdS1B and the *EcoDXXI* S subunit reveals 22% identity and 47% similarity. An alignment of the *M. pulmonis* HsdS1 peptides with the S subunits of *EcoDXXI* and *StyR124I* is shown in Fig. 7. The *M. pulmonis* peptides lack the TAEI direct repeats characteristic of S subunits of type IC enzymes; these repeats are also lacking in type IA and IB enzymes. HsdS1A has the consensus sequence [A,G]-X-X-X-X-G-K-[S,T], a motif characteristic of ATP/GTP-binding sites (Saraste *et al.*, 1990; Walker *et al.*, 1982). This motif is not found in HsdS1B or in the S subunits of the enteric bacterial enzymes.

Identification of the major ORFs flanking the *hsd1* locus

On one side of the *hsd1* locus (the left end in Fig. 3) is the *polC* gene, about two-thirds of which is present on the 6.0 and 6.6 kb fragments. This gene was identifiable because

the amino acid sequence of its predicted gene product shares 41% identity and 60% similarity with the analogous peptide from *Bacillus subtilis* (Hammond *et al.*, 1991). Because *polC* encodes the major subunit of the DNA polymerase III holoenzyme, it is assumed that the *polC* gene product plays no role in either inversion or functioning of the *hsd1* locus.

On the other side of the *hsd1* locus (the right end in Fig. 3; top of Fig. 4), two overlapping ORFs are present. The first of these ORFs would encode a peptide sharing 45% identity and 64% similarity with the amino acids of the glutamyl-tRNA synthetase (GluRS) peptide of *Bacillus stearothermophilus*, encoded by *gluX* (Breton *et al.*, 1990). Conserved in the mycoplasmal GluRS enzyme are several motifs (not shown in Fig. 4) characteristic of GluRS and other aminoacyl-tRNA synthetases from various bacterial species (Breton *et al.*, 1990; Eriani *et al.*, 1990). The ORF overlapping with *gluX*, designated ORFA, would encode a peptide containing only 126 amino acids. There was no significant similarity between the predicted amino acids encoded by ORFA and other sequences in the GenBank/EMBL databases. Presumably, because of its overlapping nature with *gluX*, a peptide encoded by ORFA would have a function related to protein metabolism and would not be involved in inversion of the *hsd1* locus. In particular, no obvious similarities between the predicted

protein encoded by ORFA and known examples of DNA invertases, or other site-specific recombinases, were noted.

DNA inversion and *hsd1* gene expression

The genes *hsdS1A*, *hsdR1* and *hsdM1* may be within a single operon because the coding regions of *hsdS1A* and *hsdR1* overlap by eight nucleotides, and only three nucleotides separate the coding regions of *hsdR1* and *hsdM1*. However, it is possible that promoters are present within coding sequences of some of the *hsd1* genes. When the invertible element is in orientation II (see Figs 3 and 4), *hsdS1A* is preceded by a Shine–Dalgarno sequence which serves as a potential ribosome-binding site (Shine and Dalgarno, 1974). Presumably, a promoter is present between the 3' end of ORFA, which is immediately followed by a putative Rho-independent transcription terminator (Platt, 1986), and the *hsdS1A* ribosome-binding site. A promoter-like sequence is present in this region (see Fig. 4), but identification of the actual promoter will require analysis of *hsdS1A* mRNA transcripts. Conversely, *hsdS1B* (when the invertible element is in orientation II) lacks a ribosome-binding site. Also, a putative Rho-independent transcription terminator is downstream from *polC* and located just 10–15 bp upstream from the *hsdS1B* initiation codon. Therefore, it is likely that *hsdS1A*, but not *hsdS1B*, is expressed when the invertible element is present in orientation II. However, when the element is in orientation I, the ribosome-binding site precedes *hsdS1B* and not *hsdS1A* and the translation terminator precedes *hsdS1A* and not *hsdS1B*. Therefore, inversion of the *M. pulmonis* *hsd1* locus should act as a genetic switch regulating expression of *hsdS1A* and *hsdS1B*.

Northern hybridization analysis has revealed that *hsd* transcripts are present regardless of the orientation of *hsd1* (data not shown), but interpretation of these data is complicated because it is not known which transcripts are derived from *hsd1* and which transcripts are derived from *hsd2*. Southern hybridization analysis and limited DNA sequence analysis of a cloned portion of *hsd2* (our unpublished data) indicate that the *hsd2* locus is structurally organized similarly to the *hsd1* locus, containing two *hsdS* genes designated *hsdS2A* and *hsdS2B*. Whether *hsd2* is another invertible element is unknown. Once the *hsd2* locus has been fully characterized, a study of the mRNA transcripts from the *hsd1* and *hsd2* loci should be easier to experimentally design and interpret.

Inversion of *hsd1* regulates R-M in *M. pulmonis*

The ability of mycoplasma virus P1 to infect cells in which

Table 1. R-M properties of *M. pulmonis*.

Virus stock used for infection	Pfu on <i>M. pulmonis</i> Strain ^a	
	16 (<i>hsd1</i> in orientation II)	52 (<i>hsd1</i> in orientation I)
P1•16	8.4×10^6 (1.0) ^b	1.2×10^7 (1.4)
P1•52	2.0×10^4 (10^{-3})	1.3×10^7 (1.0)

a. Similar results were obtained in four independent experiments. Strain KD735-16 is abbreviated as 16, and strain KD735-16-52 is abbreviated as 52.

b. Relative titre in parenthesis.

the *hsd1* locus is in either orientation was examined to determine whether DNA inversion affected the cell's R-M properties. The data obtained for strains KD735-16-52 (*hsd1* locus in orientation I) and its parent strain KD735-16 (*hsd1* locus in orientation II) are shown in Table 1. P1 virus propagated on KD735-16 (P1•16) had essentially equal titres when assayed on lawns of KD735-16 and on lawns of KD735-16-52. Conversely, P1 virus propagated on KD735-16-52 (P1•52) had a titre nearly 1000-fold lower when assayed on KD735-16 cells versus KD735-16-52 cells. These data indicate that KD735-16 cells can both restrict and modify P1 virus. In contrast, R-M of P1 virus was not detected in KD735-16-52. Essentially identical data were obtained when the R-M properties of strains KD735-16-12 and 6a14 (*hsd1* in orientation II) were compared to strains KD735-12-1 and 6a14/4 (*hsd1* in orientation I), as summarized in Fig. 1. R-M of P1 virus was detected only in strains containing *hsd1* in orientation II; strains containing *hsd1* in orientation I failed to either restrict or modify P1 virus.

Origin of the *hsd1* locus

Because type I R-M systems have previously been found only in Gram-negative bacteria and because the type IC systems to which *hsd1* is most similar are located on conjugative plasmids, it might seem that *hsd1* may have been acquired via horizontal gene transfer from a Gram-negative bacterium. However, the coding regions of the *hsd1* locus have a base composition of only 23.4 mol% G+C, typical of mycoplasma genomes (Herrmann, 1992), but significantly different from the 43.5 mol% G+C content of the R124 *hsd* locus. Moreover, the predicted amino acids encoded by the *hsd1* locus contain 16 tryptophan residues, 15 of which are encoded by TGA (see Fig. 4). TGA is the codon most often used by mycoplasmas to encode tryptophan (Muto *et al.*, 1992), but TGA is a termination codon in other eubacteria. Because the proteins encoded by the *M. pulmonis* *hsd1* locus would not even be synthesized in other bacterial systems, this locus must be mycoplasma in origin.

half of the HsdS1_B peptide is related to the entire second half of the molecule and, to a lesser degree, HsdS1_A exhibits a similar duplicative nature. Accordingly, the two halves of the HsdS1 molecule may be functionally similar. Based on analogy with other type I systems (Cowan *et al.*, 1989; Dryden *et al.*, 1993; Gann *et al.*, 1987; Gubler *et al.*, 1992; Taylor *et al.*, 1992), each half of HsdS1 probably contains a domain involved in DNA recognition and a second domain involved in interaction with HsdM1 subunits.

The particular type I enzymes to which the *M. pulmonis* system is most similar are located on conjugative plasmids, R124 and pDXXI, which may be vehicles for the dissemination of these enzymes. Although type I enzymes have heretofore been found exclusively in Enterobacteriaceae, the codon usage data presented here clearly indicate that the *M. pulmonis* system has been in a mycoplasmal genetic background for an extended period of time. Perhaps other mycoplasmas and their Gram-positive bacterial relatives also harbour type I R-M systems that have not yet been recognized. Recently, a two-subunit R-M enzyme from *Bacillus coagulans* has been described (Kong *et al.*, 1993). The nucleotide sequence (GenBank accession number L18759) of the genes encoding this enzyme reveals strong similarity with the R and M subunits of type I enzymes. Given the unusual nature of the *M. pulmonis* system, e.g. being located on an invertible element, and given some unusual features of the *B. coagulans* enzyme (Kong *et al.*, 1993), it seems that type I enzymes in Gram-positive bacteria and mycoplasmas may have evolved in interesting directions not yet seen in Gram-negative bacteria.

Infection with mycoplasma virus P1 was used to examine the R-M properties of *M. pulmonis*. Interpretation of R-M data in this system is complicated by the frequent existence of genetic differences, in addition to inversion of *hsd1*, between the strains examined. Many strains contain RFLPs unrelated to *hsd1* inversion, and some strains have changes in surface antigen make-up that may affect P1 viral infectivity (see Fig. 1). However, some pairs of strains (KD735-16 versus KD735-16-52; KD735-16-12 versus KD735-16-12-1) have no discernible genetic or antigenic differences other than their orientation of *hsd1*. A comparison of these strains indicates that inversion of *hsd1* does regulate R-M activity in this system. Of course, *M. pulmonis* may have additional, perhaps unregulated, R-M activities not encoded by *hsd1*, e.g. enzymes encoded by *hsd2*, that cannot be assayed using P1 virus; P1 DNA would be appropriately modified to be resistant to the restriction activity of any R-M system expressed in all strains.

In general, site-specific DNA inversions are catalysed by recombinases that are encoded by a gene that is located either within or immediately adjacent to the invertible

element (Dybvig, 1993). The *hsd1* locus is apparently an exception to this rule. None of the ORFs located in the vicinity of this locus is a likely candidate for a recombinase-encoding gene. Recently, we have determined a partial nucleotide sequence of the *hsdS2_B* gene, so called because it is located downstream from and on the complementary strand to the *hsdM2* gene (our unpublished data). The nucleotide sequence of the 5' half of this gene is identical to the sequence of the 5' halves of *hsdS1_A* and *hsdS1_B*. If the *hsd2* locus is another invertible element, perhaps the same recombinase promotes inversion of both *hsd1* and *hsd2* via recognition of the conserved sequences in the *hsdS* genes. If so, the gene encoding the recombinase may be located in the vicinity of *hsd2* rather than *hsd1*.

Why does *M. pulmonis* regulate an R-M system via DNA inversion, and what is the fate of the *M. pulmonis* chromosome when *hsdS1_A* is switched from the off to the on orientation? When in orientation I, *hsdS1_A* is presumably not expressed and chromosomal DNA would not be modified at HsdS1_A recognition sequences. When switched to orientation II, does HsdS1_A-directed restriction activity result in double-strand breaks in the chromosome? If not, some factor may protect the chromosome from nuclease activity. Several factors have been described that alleviate restriction activity in a variety of systems (Bickle and Kruger, 1993). If double-strand breaks are generated, they may sometimes be lethal, perhaps explaining why subclones in which *hsd1* has switched from orientation I to orientation II were isolated relatively rarely. In some cells, these breaks may be repaired, perhaps resulting in chromosomal rearrangements (Dybvig, 1993). Every subclone examined to date in which *hsd1* has inverted from orientation I to orientation II also contains an RFLP involving recombination between precursor fragments of 4.9 and 5.7 kb (see Fig. 1). In addition, recombination between the 4.9 and 5.7 kb fragments has not been observed in the absence of concurrent inversion of *hsd1*. The basis for this apparent correlation is unknown. The 4.9 and 5.7 kb fragments contain a cluster of genes encoding phase-variable surface antigens (Bhugra, 1992). Whether DNA rearrangements in this region regulate surface antigen expression remains to be examined.

Experimental procedures

Mycoplasmas and viruses

All mycoplasmas used in this study have been previously described and are derivatives of *M. pulmonis* strain UAB 6510 (Dybvig *et al.*, 1988; Bhugra and Dybvig, 1992). Mycoplasmas were propagated in mycoplasma medium as previously described (Dybvig and Cassell, 1987). Mycoplasma virus P1 was assayed as pfu on lawns of susceptible cells as described by Dybvig *et al.* (1987).

Molecular cloning

Mycoplasmal chromosomal DNA was isolated as described by Dybvig and Alderete (1988). *Hind*III fragments of mycoplasma chromosomal DNA were purified from agarose gels and cloned into the *Hind*III site of plasmid pUC18 using standard techniques (Sambrook *et al.*, 1989). Conditions for enzymatic amplification using PCR were as described previously (Dybvig and Woodard, 1992), and PCR products were cloned into the pCR-II cloning vector (Invitrogen). The PCR product spanning the junction between the 6.2 and 6.6 kb fragments was amplified using oligonucleotides 5'-GTCATCGTTCTAATGCTGG-3' (nucleotides 5542–5560 in Fig. 4) and 5'-GTTAATCTAT-TATTTGGGTATTTAC-3' (complementary strand of nucleotides 6825–6802 in Fig. 4). Amplification of *hsdS1b* from KD735 was accomplished using oligonucleotides 5'-CTTTAAACAGAAAAAGTGCTGTC-3' (nucleotides 2998–3020 in Fig. 4) and 5'-AAAAATGAATATTCATTATCTAT-GAG-3' (nucleotides 8712–8737 in Fig. 4, but located on the strand complementary to the sequence shown in Fig. 4 when the *hsd1* locus is present in orientation I, as is the case for KD735 DNA).

Hybridization analysis

DNA fragments were resolved on 0.8% agarose gels. Mycoplasma RNA was isolated by phenol–chloroform extraction of acid guanidinium thiocyanate lysates (Chomczynski and Sacchi, 1987) and resolved on 1.0% formaldehyde gels. Nucleic acids were transferred to nitrocellulose and hybridized with probes labelled with ³²P by nick translation. Normal (high)-stringency conditions were used for all hybridization experiments. Conditions for electrophoresis, membrane blotting, radiolabelling of probes, and hybridization were as described (Sambrook *et al.*, 1989).

DNA sequencing and sequence analysis

DNA sequencing was performed on double-stranded plasmid templates by the dideoxy nucleotide chain termination method with Sequenase version 2.0 (US Biochemical). DNA oligonucleotide primers for sequencing and PCR were supplied by the Oligonucleotide Synthesis Core Facility of the University of Alabama at Birmingham, USA. In all cases, the nucleotide sequences of both strands of DNA were determined to ensure accuracy. Computer analysis was performed using the Genetics Computer Group (GCG) programs (University of Wisconsin, Madison) on a VAX-VMS computer at the University of Alabama Cancer Center. Initial comparisons between *M. pulmonis* proteins and the GenBank/EMBL databases were performed using the programs *TFASTA* and *BLAST*. The *GAP* program was used to obtain calculations of per cent identity and per cent similarity between aligned sequences. Multiple sequence alignments were obtained using the program *PILEUP*.

Acknowledgements

We thank Portia Bowen, Ramakrishnan Sitaraman and Ann Woodard for technical assistance. This research was supported by Public Health Service Grant AI31144 from the

National Institutes of Health. The Genetics Computer Group programs were made available to us through the University of Alabama at Birmingham (UAB) Center for AIDS Research, supported by PHS Grant P30 AI27767. Synthesis of DNA primers was supported by NCI Grant CA13148 to the UAB Comprehensive Cancer Center.

References

- Bhugra, B. (1992) Analysis of high-frequency chromosomal rearrangements in *Mycoplasma pulmonis*. Ph.D. Thesis, University of Alabama at Birmingham, Birmingham, Alabama, USA.
- Bhugra, B., and Dybvig, K. (1992) High-frequency rearrangements in the chromosome of *Mycoplasma pulmonis* correlate with phenotypic switching. *Mol Microbiol* **6**: 1149–1154.
- Bhugra, B., and Dybvig, K. (1993) Identification and characterization of IS1138, a transposable element from *Mycoplasma pulmonis* that belongs to the IS3 family. *Mol Microbiol* **7**: 577–584.
- Bickle, T.A., and Kruger, D.H. (1993) Biology of DNA restriction. *Microbiol Rev* **57**: 434–450.
- Breton, R., Watson, D., Yaguchi, M., and Lapointe, J. (1990) Glutamyl-tRNA synthetases of *Bacillus subtilis* 168T and of *Bacillus stearothermophilus*. *J Biol Chem* **265**: 18248–18255.
- Chomczynski, P., and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **122**: 156–159.
- Cowan, G.M., Gann, A.A.F., and Murray, N.E. (1989) Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* **56**: 103–109.
- Doolittle, R.F., Feng, D.F., Johnson, M.S., and McClure, M.A. (1986) Relationships of human protein sequences to those of other organisms. *Cold Spring Harbor Symp Quant Biol* **LI**: 447–455.
- Dryden, D.T.F., Cooper, L.P., and Murray, N.E. (1993) Purification and characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J Biol Chem* **268**: 13228–13236.
- Dybvig, K. (1993) DNA rearrangements and phenotypic switching in prokaryotes. *Mol Microbiol* **10**: 465–471.
- Dybvig, K., and Alderete, J. (1988) Transformation of *Mycoplasma pulmonis* and *Mycoplasma hyorhinis*: transposition of Tn916 and formation of cointegrate structures. *Plasmid* **20**: 33–41.
- Dybvig, K., and Cassell, G.H. (1987) Transposition of gram-positive transposon Tn916 in *Acholeplasma laidlawii* and *Mycoplasma pulmonis*. *Science* **235**: 1392–1394.
- Dybvig, K., and Woodard, A. (1992) Cloning and DNA sequence of a mycoplasma *recA* gene. *J Bacteriol* **174**: 778–784.
- Dybvig, K., Liss, A., Alderete, J., Cole, R.M., and Cassell, G.H. (1987) Isolation of a virus from *Mycoplasma pulmonis*. *Isr J Med Sci* **23**: 418–422.
- Dybvig, K., Alderete, J., Watson, H.L., and Cassell, G.H. (1988) Adsorption of *Mycoplasma pulmonis* virus P1 to host cells. *J Bacteriol* **170**: 4373–4375.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990) Partition of tRNA synthetases into two classes

- based on mutually exclusive sets of sequence motifs. *Nature* **347**: 203–206.
- Gann, A.A.F., Campbell, A.J.B., Collins, J.F., Coulson, A.F.W., and Murray, N.E. (1987) Reassortment of DNA recognition domains and the evolution of new specificities. *Mol Microbiol* **1**: 13–22.
- Gorbalenya, A.E., and Koonin, E.V. (1991) Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Letts* **291**: 277–281.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., and Blinov, V.M. (1989) Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucl Acids Res* **17**: 4713–4730.
- Gubler, M., Bragaglia, D., Meyer, J., Piekarowicz, A., and Bickle, T.A. (1992) Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J* **11**: 233–240.
- Hammond, R.A., Barnes, M.H., Mack, S.L., Mitchener, J.A., and Brown, N.C. (1991) *Bacillus subtilis* DNA polymerase III: complete sequence, overexpression, and characterization of the *polC* gene. *Gene* **98**: 29–36.
- Herrmann, R. (1992) Genome structure and organization. In *Mycoplasmas: Molecular Biology and Pathogenesis*. Maniloff, J., McElhaney, R.N., Finch, L.R., and Baseman, J.B. (eds). Washington, D.C.: American Society for Microbiology, pp. 157–168.
- Kong, H., Morgan, R.D., Maunus, R.E., and Schildkraut, I. (1993) A unique restriction endonuclease, *Bcgl*, from *Bacillus coagulans*. *Nucl Acids Res* **21**: 987–991.
- Lauster, R. (1989) Evolution of type II DNA methyltransferases: a gene duplication model. *J Mol Biol* **206**: 313–321.
- Murray, N.E., Daniel, A.S., Cowan, G.M., and Sharp, P.M. (1993) Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol Microbiol* **9**: 133–143.
- Muto, A., Andachi, Y., Yamao, F., Tanaka, R., and Osawa, S. (1992) Transcription and translation. In *Mycoplasmas: Molecular Biology and Pathogenesis*. Maniloff, J., McElhaney, R.N., Finch, L.R., and Baseman, J.B. (eds). Washington, D.C.: American Society for Microbiology, pp. 331–347.
- Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annu Rev Biochem* **55**: 339–372.
- Price, C., Lingner, J., Bickle, T.A., Firman, K., and Glover, S.W. (1989) Basis for changes in DNA recognition by the *EcoR124* and *EcoR124/3* type I DNA restriction and modification enzymes. *J Mol Biol* **205**: 115–125.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990) The P-loop — a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* **15**: 430–434.
- Shine, J., and Dalgarno, L. (1974) The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc Natl Acad Sci USA* **71**: 1342–1346.
- Taylor, I., Patel, J., Firman, K., and Kneale, G. (1992) Purification and biochemical characterization of the *EcoR124* type I modification methylase. *Nucl Acids Res* **20**: 179–186.
- Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982) Distantly related sequences in the α - and β -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J* **1**: 945–951.
- Wise, K.S., Yogev, D., and Rosengarten, R. (1992) Antigenic variation. In *Mycoplasmas: Molecular Biology and Pathogenesis*. Maniloff, J., McElhaney, R.N., Finch, L.R., and Baseman, J.B. (eds). Washington, D.C.: American Society for Microbiology, pp. 473–489.
- Yogev, D., Rosengarten, R., Watson-McKown, R., and Wise, K.S. (1991) Molecular basis of *Mycoplasma hyorhinis* surface antigen variation: a novel set of divergent genes undergo spontaneous mutation of periodic coding regions and 5' regulatory sequences. *EMBO J* **10**: 4069–4079.